DISCRETE BRAND CHOICE MODELS: ANALYSIS AND APPLICATIONS

A Thesis
Presented to
The Academic Faculty

by

Liyu Zhu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
August 2007

DISCRETE BRAND CHOICE MODELS: ANALYSIS AND APPLICATIONS

Approved by:

Professor Augustine Esogbue, Committee Chair
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Paul Griffin
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Jye-Chyi (JC) Lu
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor MinQiang Li
College of Management
Georgia Institute of Technology

Professor Patrick McCarthy
College of Economics
Georgia Institute of Technology

Date Approved: July 2007

To My Parents

ACKNOWLEDGEMENTS

I am very grateful to my advisor, Professor Augustine Esogbue, for his unselfish help and for his providing me with a very good computational environment in his Intelligent Systems and Controls Lab which considerably facilitated my research. I also appreciate his role in organizing, reviewing, and making the presentation of this dissertation clearer.

I would like to thank my committee members: Professors Paul Griffin, Jye-Chyi Lu, MinQiang Li, and Patrick McCarthy for their help; Professor Paul Griffin for his persistent support, and Professor Jye-Chyi Lu for his excellent suggestions.

I am in debted to the Dean's office of the College of Engineering at Georgia Tech, which has provided financial support for my research, making it possible for me to finish my dissertation. My special thanks go to Dr.Gail DiSabatino, for her kind help and overall support. She was one of the best and most valuable friends of Georgia Tech students.

I am very grateful to Professor Eric Carlen in the School of Mathematics at Georgia Tech for his consistent help and support. I like to thank Professor Greg Allenby at Ohio State University for his discussion in Chapter 3. I also appreciate Professor Floyd Hanson in the University of Illinois at Chicago for often answering my questions in Chapter 4.

I want to express my appreciation to my friends for their friendship and for playing sports together.

Finally, my parents' unconditional support, patience, and love are in my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS OR ABBREVIATIONS

Symbol definitions or abbreviations in chapter two.

$V_1^*(x)$:	discount maximum cumulative profit of company A from t_0 to ∞
$V_2^*(x)$:	discount maximum cumulative profit of company B from t_0 to ∞
$x(t)$, $y(t)$:	share at time t
$u(t)$:	advertising expenditure rate (effort)
r:	discount coefficient
m_1:	profit coefficients of company A
m_2:	profit coefficients of company B
c_1:	cost coefficients of company A
c_2:	cost coefficients of company B
ρ_1:	advertisement response coefficients for company A
ρ_2:	advertisement response coefficients for company B
HJB:	Hamiliton-Jacobi-Bellman equation
SDJG:	Stochastic differential-Jump Game
H_i :	Hamilton function of player i
p or p_k	transition probability of markov chain
λ_J	jump rate
$\mathcal{J}_i$	integrant of jump in HJB of player i
$dP(t)$:	Mark-time Poisson process
$\mathcal{P}(t)$:	measure for mark-time Poisson process
h:	discrete step length in x dirction
n_w:	number of Browian motion
n_p:	number of Mark-time Poisson process
σ:	volatility of diffusion process
$V(u)$:	abbreviation of $V(x,u)$

Symbol definitions or abbreviations in chapter three.

y_{ij} :	latent variable of profile j valuated by the respondent i
x_{ij} :	variables associated with profile j for respondent i
β_i :	respondent i-specific coefficients, $\beta_i = (\beta_{i1}, ..., \beta_{ik})'$
ε_{ij} :	i.i.d norm error term $Norm(0, \sigma^2)$
Γ :	matrix that relates β_i to the value of z_i
Σ :	covariance matrix
z_i :	customer(respondent)'s information
ξ_i :	unobserved heterogeneity component
w_i :	the weight, Gamma distribution $\mathcal{G}amma(\nu_1/2, \nu_1/2)$
MCMC:	Markov Chain Monte Carlo Simuylation
MLE:	Maximum Likelihood Estimation
AIC:	Akaike Information Critiera
BIC:	Bayesian Information Critiera
LMD:	Log Margin Density
L:	Likelihood function
L_h, L_r, L_w :	log likelihood function of top level, robust regression level, and overall model
Φ	cumulative probability function of normal distribution
t	t distribution
ν	freedom of t distribution
$\sim$	distribution
$\propto$	proportional to
$'$	transpose of a vector or a matrix
π	probability density

Symbol definitions or abbreviations in chapter four.

Q_k:	products in the time period k
$R_k(D_k)$:	revenue function of D_k
c_k:	production cost incurred in period k
h_k:	inventory cost from period $k-1$ to k
q^*:	dynamic price at each time
X_k :	production at discrete time k
I_k :	inventory at discrete time k
$\lambda(t)$:	the demand rate
$I(t)$:	inventory rate
$X(t)$:	production rate
$h(t)$:	inventory cost in unit time
$c(t)$:	production cost in unit time
$q(t)$:	price
$Q(t)$:	capacity
ϕ:	demand at each arrival
σ:	volatility
α, θ:	constants of CIR
$W(t)$:	Browian motion
$dP(t)$:	Mark-time process
$p_k\{(\lambda, X), (\lambda, X)\}$:	probability at time k from (λ, X) to (λ, X)
$p_k\{(\lambda, X), (\lambda \pm h_\lambda, X)\}$:	probability at time k from (λ, X) to $(\lambda \pm h_\lambda, X)$
z:	jump amplitude value
λ_J:	jump rate
T:	the upper bound of time period
$V^*(X, \lambda, t)$:	cumulative profit from t_0 to t
$V_k^*(X, \lambda)$:	cumulative profit from t_0 to t_k

SUMMARY

Discrete brand choice is a microeconomics problem that is concerned with demand predictions, pricing, and how personal choice behaviors affect the supply-demand equation. It is an important problem addressed by many, including McFadden and Heckman, who won the Nobel Prize in 2000 for their work on this topic. The discrete brand choice problem involves“selection among alternative sets in markets to maximize a customer's own self-interest defined by a utility function under the consumer consumption level constraint.” The role of models, including those of the operations research variety, in advancing the state of the art of this problem domain has been postulated. However, very few discrete brand choice models encountered in the literature study the choice dynamics from the market share perspective. There is clearly a need for more precise integration of more information and robust estimation techniques.

In this thesis, we study the brand choice problem from the following three perspectives: a company's market share management, introduction of customers with different perspectives, and an analysis of an application domain that is illustrative of these issues. Our contributions following these perspectives include (1) the development of a stochastic differential-jump game (SDJG) model for brand competition in a specific situation wherein market share is modeled by a jump-diffusion process, (2) a robust hierarchical logit/probit model for market heterogeneity, and (3) applications of a logit/probit model to the dynamic pricing problem occurring in production-inventory systems with jump events.

First, we develop an SDJG model for brand competition. An SDJG model has the ability to model continuous variables and jump/discrete events simultaneously. Jump events are modeled by mark-time Poisson processes, with advertising effort as the control input. The Hamilton-Jacobi-Bellman (HJB) equation is the modeling framework proposed for this problem, but we resort to a Markov chain approximation method as the favored computational strategy because of the difficulty of obtaining a closed-form solution for a general HJB. The SDJG model is explicitly defined and applied in market competition. To our knowledge, this is the first application of this type in the literature.

Next, we consider a Bayesian robust hierarchical logit/probit model in the analysis of market heterogeneity. A hierarchical model combines features of products with characteristics of customers. Markov Chain Monte Carlo (MCMC) simulation is used for the estimation component. A robust hierarchical logit/probit model is then developed and validated using credit card data from a regional bank system. Our model represents an improvement over a general hierarchical logit/probit model based on better prediction precision and higher log margin density (LMD).

Finally, we employ a logit/probit model by imbedding it in a dynamic pricing problem. The resultant dynamic pricing model integrates information of production, inventory, and customers' choice. We not only consider more complex demand processes modeled by the Cox-Ingersoll-Ross (CIR) process, but also production systems with jump events. The dynamic pricing problem considered here is a more complex and comprehensive situation than those found in current literature.

Our research explores the use of a quantitative method of operations research to control the dynamics of market share and provides a precise estimation method to integrate more detailed information in discrete brand choice models.

Keywords: **Discrete Brand Choice, Stochastic Differential-Jump Game, Markov Decision Process, Robust Hierarchical Bayesian Logit/Probit, Dynamic Pricing**.

CHAPTER I

INTRODUCTION

1.1 Research Problem Statement

Brand choice is the foundational topic of central interest in this dissertation. It is the subject of various constituencies, including organizations represented by companies, firms, etc., and their customers. It is instructive to consider the perspective of these groups with respect to the concept of brand choice.

For a customer, brand choice is deciding which type of products, services, and plans to select from some alternative sets in a market so as to maximize his/her own self-interest, which is defined by John Hicks and Paul Samuelson (Samuelson 1970) as stable and innate preferences. This self-interest can be represented as a utility function $U(x)$, where x is the consumer consumption level. A customer wishes to maximize the utility function subject to constraints such as budget or availability. The problem of brand choice is of central interest to both customers and researchers. An array of models with varying degrees of sophistication, has been proposed in the literature for studying the problem of brand choice. Although the well-known logit and probit models (McFadden 1974, 2001) may be simple, a customer's decision is actually quite complex because of many factors such as culture, demographics, economics, physiology, etc. Marketing researchers endeavor to understand, describe, and predict consumer choices such as those made for a particular brand, service, or store. For example, customers may need to choose where to shop, when to buy a product, and which brand to purchase. On the company side, firms also may need to think about what type of customer to target, how to cross-sell and retain customers, what price to charge, how to assess the effectiveness of promotions and advertising, etc. In recent years, these questions have become important in customer relationship management (CRM).

A basis for brand choice research is deeply rooted in operations research techniques such as statistics, mathematical modeling, optimization, and simulation. The statistical techniques most commonly used are maximum likelihood estimation, Bayesian theory, and statistics tests. The most popular simulation method used is the Markov Chain Monte Carlo simulation. These statistical methods are used on a more detailed level to estimate parameters. Pertinent optimization techniques such as dynamic programming, can be expressed as a Markov decision process (MDP) and used to model the dynamic brand choice process. At a market share management level, game models are used for getting the best strategy such as advertisement effort. Nonlinear programming is embedded in the maximum likelihood estimation. Brand choice actually is a difficult problem that needs to be addressed. The reasons for the difficulty are discussed in detail in Section 1.1.3.

To motivate our interest, we note that Daniel McFadden of the University of California at Berkeley shared the 2000 Nobel Prize in economic science with James Heckman of the University of Chicago because of their theoretical analysis of discrete choices. Others who have made important contributions in this area include Griliches (1957), Thurstone (1927), Marschak (1960), Luce (1959), Tversky and Kahneman (1974,1981), Ben-Akiva and Morikawa (1990), Manski (1977), and Train (1998).

The current research on brand choice has many applications, some of which are transportation, health care, production development, credit card profile design, insurance plans, the automobile industry, household production, electronic products, etc.

1.1.1 Application Areas

In this section, some important application areas are listed to show the practical relevance of this research topic.

Transportation. In the 1970s, Peter Diamond and Robert Hall at MIT developed a separable-utility, multi-stage budgeting, representative consumer model for the complexity of consumer transportation decisions, including frequency, timing, and destination of shopping trips. Since the 1970s, some researchers at UC Berkely, Mcfadden, Train, and Manski, have been attracted to this problem domain and subsequently developed useful tools for transportation planning based on microeconomic analysis of individual decisions. Also at MIT, Hauser, Koppelman, and Tybout (1981) have also done research about transportation and facility location based on discrete choice models.

Health Care. Health care is expensive for almost everyone, especially in the United Sates. Cardon and Hendel (2001) structurally estimated a model of health insurance and health care choices using data on single individuals. They tested for unobservable links between health insurance status and health care consumption. Heckman (1979) has also done some research on health care provider choice based on discrete choice models. Ryan and Gerard (2003) provided a good review about discrete choice as it relates to health care economics.

Banking and Insurance. Allenby and Ginter (1995) used a Bayesian hierarchical model to study credit card profile choices. Their data was from an experiment in which partial profiles of credit cards were presented to 946 respondents under 14,799 observations. Greene (1998) examined three models for sample selection that are relevant to modeling credit scoring by commercial banks. They used a binary choice model to examine the bank's decision of whether or not to extend credit. Stango (2000) have also studied credit card issuers with "fixed rate" and "variable rate."

Automobile Industry. McCarthy et al. (1992) studied the loyalty and the switch rate of automobile brands. Berry et al. (1995) developed techniques for empirically analyzing demand and supply in product markets and then applied these techniques to analyze equilibrium in the U.S. automobile industry.

Household Products. There is a great deal of research on household product brand choices such items as toilet paper, ketchup, soup, and laundry detergent. Kamakura et al. (1996) studied the preference heterogeneity in the peanut butter market using a nested logit model. Gonul and Srinivasan (1996) also studied the effect of coupons on purchase behavior of disposable diapers, using a stochastic programming model (MDP). Erdem and Keane (1996) used decision making under uncertainty technology to capture dynamic brand choice processes in turbulent consumer goods markets (liquid detergent). Erdem et al. assumed a customer choice of a brand according to dynamic programming to maximize the utility function. Dillon and Gupta (1996) used a segment-level model to study and categorize volume and brand choice at the same time by employing a nested logit model.

The current literature contains an array of models for brand choice. The most popular of these models including logit, nested logit, probit, Bayesian methods, structural models, and dynamic models, will be reviewed in the next section.

1.1.2 Overview of Model Methodology

In this section, we review major methodologies described in the literature for modeling brand choice. All models are categorized into two classes: (1) a static model, and (2) a dynamic model. A static model is akin to a picture of customers' choices where there are no time and space changes. A dynamic model is like a video of customers' choice processes where there are time and space changes. We begin with the logit/probit model.

(1) **Static Models**. The type of probability models of interest here include the logit model, nested logit model, and probit model.

(a) **Logit Model**

The logit model computes the probability of choosing a brand as a function of the attributes of all brands available. Hlavac and Little (1969) used a somewhat similar model to the logit model to calculate the probability that an automobile buyer purchases a car at a particular dealership. Silk and Urban (1978) imbedded a logit model in their pre-test-market evaluation process for new products. Punj and Staelin (1978) applied this model to students choosing a business school. Gensch and Recker (1979) compared the fitting ability of a logit model with that of regression for shoppers choosing grocery stores.

Marschak (1960) showed that those following axioms imply that the model is consistent with utility maximization. The relation of the logit formula to the distribution of unobserved utility (as opposed to the characteristics of choice probabilities) was developed by Marley, as cited by Luce et al. (1965). McFadden (1974) showed that the logit formula for the choice probabilities necessarily implies that unobserved utility is an extreme value. In his Nobel lecture, McFadden (2001) provided a fascinating historical review of the development of this model.

Assumed axioms: Consumers must choose one from a brands set S that contains different brands. Other assumptions include

1. Brand $k \in S$ has a utility function:

$$U_k = V_k + \epsilon_k$$

   V_k and ϵ_k are the deterministic and random components of utility function, respectively.

2. Among those brands, the customer chooses a brand with the maximum utility function, i.e., The brand i is chosen because

$$p_i = P\{U_i \geq U_k, \quad k \in S\}$$

3. The random component of a utility function is the double exponential random variable (extreme value variable).

$$P\{\epsilon_k < \varepsilon\} = e^{-e^{-\varepsilon}}$$

From the foregoing assumptions, Theil (1971) and McFadden (1974) derived the probability that brand i was chosen as

$$p_i = \frac{e^{V_i}}{\sum_{k \in S} e^{V_k}}$$

(b) **Nested Logit**

A nested logit model is suitable when the set of brands can be partitioned or grouped into subsets, called "nests." In fact, this is a hierarchical (multi-level) logit model. Wen and Koppelman (2001) derived various cross-nested models as special cases of the general nested logit (GNL).

(c) **Probit**

Marschak (1960) translated economic terms into a utility. Hausman and Wise (1978), and Daganzo (1979) elucidated the generality of the specification for representing various aspects of choice behavior. Utility is decomposed into observed and unobserved parts V_j, ϵ_j, respectively; V_j and ϵ_j are defined as the same as the above axioms (1).

An example of a probit model is exhibited in the following. Consider the vector composed of each $\epsilon_j = \epsilon_{j1}, ..., \epsilon_{jm}$. We assume that ϵ_j is distributed normally with a mean vector of zero and a covariance matrix Ω. The density of ϵ_j is

$$\phi(\epsilon_j) = \frac{1}{\sqrt{2\pi}^m |\Omega|^{\frac{1}{2}}} e^{-\frac{1}{2}\epsilon_j^{'}\Omega^{-1}\epsilon_j}$$

The covariance can depend on the variables faced by a customer. $'$ is the transpose of a vector. The choice probability is

$$p_i = P(V_{ji} + \epsilon_{ji} > V_{jk} + \epsilon_{jk}, \quad k \neq i) = \int P(V_{ji} + \epsilon_{ji} > V_{jk} + \epsilon_{jk}, \quad k \neq i)\phi(\epsilon_j)d\epsilon_j$$

where the integral is over all values of ϵ_j.

(d) **Structural Models**

Structural models are decomposed into two main classes: structural equation models and hierarchical models. These models mirror the analysis and observation processes of human beings.

A structural equation model (SEM) is a general statistical modeling technique to establish relationships among variables. This type of model is more often used in customer physiology research. Baumgartner and Homburg (1996) provided a very good review of SEM. It appears that the hierarchical model has become more popular in marketing research because it is a structural model that matches the structure of marketing data and MCMC technologies.

In a hierarchical model, a hierarchy describes a model that consists of units grouped into different levels. In these models, there are a number of lower level units within each higher level unit. Especially in the case of the logit or probit model with a binary (0,1) response, this is no longer adequate. Results by Rodriguez and Goldman (1995) illustrated this problem. Goldstein and Rasbash (1996) described this model, too.

In the sequel, we present an overview of a three-level logit model: A three-level logit model is a two-level logit model with a lower level regression for each β_i, the parameter associates choice attributes.

First level: Here, observation y_i indicates if brand i is chosen. y_i as a Bernoulli random variable with success probability $y_i|p_i$, $y_i = 1$ with probability p_i. $y_i = 1$ means brand i is chosen. $y_i = 0$ means brand i is not chosen.

The second level: It is a logit model. $x_i = (x_{i1}, ..., x_{ik})'$ are the choice attributes; $\beta_i = (\beta_{i1}, ..., \beta_{ik})'$ are the parameters associated with these choice attributes.

$$p_i = \frac{\exp(x_i'\beta_i)}{\sum_j \exp(x_j'\beta_j)}$$

The third level: A linear regression is used.

$$\beta_i = \Gamma z_i' + \varepsilon$$

where $z_i = (z_{i1}, ..., z_{im})$, $\varepsilon \sim N(0, \sigma^2)$. The three-level probit model is a two-level probit model with a regression or distribution for β_i. The third level is a linear regression for β_i, where $\beta_i = \Gamma z_i' + \varepsilon$.

McCulloch and Rossi (1994) used a hierarchical probit model. Allenby and Ginter (1995) employed a hierarchical logit model to study heterogeneity. Subsequently, Yang and Allenby (2003) used a hierarchical auto regression model to address the social connection.

(2) **Dynamic Models**

The purpose of a dynamic model is to track the customer choice process, to observe brand choice over various time periods.

Rust (1987) first formulated a simple regenerative model for bus engine replacement. His model described the behavior of the maintenance plan at the Madison Metropolitan bus company.

Erdem and Keane (1996) combined the Monte Carlo, maximum likelihood estimator, and dynamic programming methods to study the customer forward-looking brand choice process. Their application focused on the household production market. Erdem et al. (2004) also considered quantity and brand choice together using a dynamic programming model to calculate the utility function.

Gonul and Shi (1998) used the estimable structural dynamic programming model to maximize direct mailer profit. Their model combines the customers' behavior and company's decisions.

Imai et al. (2005) also extended the dynamic programming framework with the Bayesian Markov Chain Monte Carlo algorithm to solve the resultant dynamic programming problem and estimate the parameters conjunctively.

Considering the importance of dynamic choice models, our research examines the brand choice problem as it may occur at the strategic management level. We propose a stochastic differential-jump game model as an appropriate framework. This is discussed in detail in Chapter 2.

1.1.3 Complexity of Brand Choice

It is apparent that brand choice decision involves a considerable array of factors. For example, the reason that a customer may choose a brand is affected by many market elements such as objective (motivating) environments, coupon, price improvement, demographics, internet access, individual habits, etc. Some of these elements that complicate the brand choice problem will now be brifly examined in the following section.

1. Motivating Environments. A study of this factor is useful for adding independent variables that reflect the personal and environmental conditions existing in everyday life, for which goods or services are used.

Some authors have included "situational" variables, operationalized as variation in activity, or in a type of objective environment for activity (e.g., Belk (1975), Ratneshwar and Shocker (1991), Miller and Ginter (1979), Dickson (1982)).

Yang et al. (2002) also studied the effects of objective environment using data collected by the Consumer Insights Group at the Miller Brewing Company, a major U.S. beer producer. They found that 1) across individuals the objective environment is associated with heterogeneous, not homogeneous, motivating conditions; 2) within an individual, motivating conditions may change with variation in the objective environment; and 3) motivating conditions are related to preferences for specific attributes.

2. Online/Offline. Internet marketing may change customer choice behavior and habit. Andrews and Currim (2000), Bakos and Brynjolfsson (2000), Brynjolfsson and Smith (2000), Burke et al. (1992), Degeratu et al. (2000), and Shankar et al. (2003), conducted empirical research exploring consumer behavior in an online environment.

Andrews and Currim (2000) found that the brand loyalty coefficient in a multinomial logit model is lower for online versus offline grocery shopping. However, in the study, online shoppers tended to select from a smaller consideration set of brands, thereby remaining loyal to a smaller number of brands.

3. Price Promotion. Price promotions are short-term cuts offered to consumers. Many researchers such as Guadagni and Little (1983) found the customer's loyalty relates to price promotion. Rju et al. (1990) pointed out that theoretical analysis shows that if all brands in the product market have high brand loyalty, none of the companies will find it profitable to use price promotions. Furthermore, a brand's likelihood of using pricing promotions increases with the number of competing brands in the product market. Rju et al. (1990) studied brand loyalty using grocery store data.

4. Coupon. Feinberg et al. (1992) and Morrison et al. (1971) derived the long-run brand choice probability for couponed brands in an equilibrium market using a Markov process model. Dhar et al. (1996) conducted experiments in the UCLA Student Union focusing on soft drinks and yogurt. They considered three different kinds of coupons using the probit model. Their results showed that coupons can increase a brand's share significantly, but may not increase profit very much.

5. Demography. Allenby and Rossi (1991) and Gupta et al. (1994), introduced demographic variables in brand choice models. Kalyanam et al. (1997) incorporated demographic variables in brand choice using data for ketchup, coffee, liquid detergent, and laundry detergent. They found the following: (1) There is a negative relationship between household income and its price sensitivity (2) the presence of children will increase the probability of choosing large size packaged products in the same category.

6. Habit. Brand choice is affected by the previous experience of customers in two ways. One way is loyalty and the other is habit. Heckman (1981,1991) call this the phenomenon habit persistence. Creating a dynamic model that includes past, present, and future brand choices is very difficult. Roy et al. (1996) developed a general dynamic brand choice model that, in fact, is a logit model. They did research on ketchup data in a Springfield, Missouri, market. The model consists of heterogeneity, state dependence, and habit (loyalty).

1.2 Research Motivation and Objectives

Our research focuses on the dynamic nature of the brand choice process, models for the choice difference among customers (i.e., heterogeneity), and applications in production-inventory systems. In this dissertation, we model different aspects of brand choice using an assortment of approaches.

This research is divided into three components. First, from a high market management level, the accumulation of brand choice of individuals is expressed as the market share at the management level. For example, a company may wish to design the best advertising policy in order to attract more customers to select its own brands. In chapter two, a stochastic differential-jump game is used to model this situation. In the second component of this research, we observe that each customer is different even though all reasonable customers are assumed to make decisions in a manner that mirrors the fundamental dynamic programming principle. We model the differences among customers using a robust hierarchical logit/probit model. This is treated in Chapter 3. Because of the many parameters to be estimated, Markov Chain Monte Carlo simulation (MCMC) is proposed as a suitable method. The third component, namely, application is discussed in Chapter 4. The brand choice model now has a new important application area, namely, dynamic pricing. We use the logit/probit as the customer's response in a manufacturing-inventory-marketing system, which integrates information of manufacturing, inventory, and marketing together. In following sections, we present an overview of each research topic and highlight their methodologies as well as their resultant contributions to the thesis topic.

In Figure (1), the dissertation layout and each component of the thesis research are displayed. The brand choice is an interaction process between companies and customers. Each different customer has his/her own profile. A company has to consider the market segmentation and maximizing its revenue. These issues are inputs of the brand choice problem. We explain the definition, application areas, and complexity of brand choice in Chapter 1. The core of this problem is the dynamic interaction

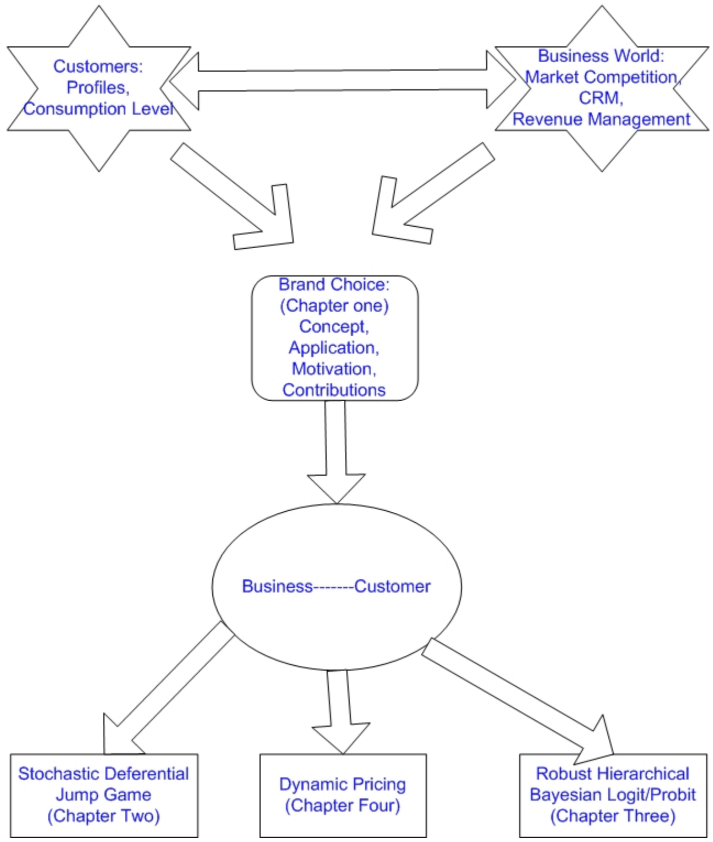

Figure 1: Overview of research structure.

between companies and customers. In Figure (1) business represents companies. Further, we address three subtopics: stochastic differential-jump game model for market competition, heterogeneity of customers, and dynamic pricing application.

1.2.1 A Stochastic Differential-Jump Game Model for Brand Competition

In Chapter 2, we consider the problem of modeling the brand competition problem via some novel approaches. This is considered a significant departure from classical approaches, with the potential for useful contributions to the brand competition literature. We begin by providing some concepts of a stochastic differential-jump game. A stochastic differential-jump game is an extension of a stochastic differential game (Basar and Olsder (1995)) that belongs to a subset of the general class of dynamic games originating from systems theory, optimal control, and dynamic programming. One such game includes a number of state variables that describe the states of a dynamic system. In a stochastic differential game, the evolution of the state variables is described by a set of diffusion equations. In the stochastic differential-jump game, however, a mark-time Poisson process (Hanson (2006)) is added for each state variable equation. Historically, differential game theories have been studied extensively by Isaacs, Friedman, Leitmann, Krasovskii, Subbotin, Basar, and Olsder.

Moorthy (1985, 1993) and Rao (1990) surveyed game theory modeling applications in marketing. Jørgensen (1986), Rao (1990), and Moorthy (1993) discussed pricing strategies. Dolan et al. (1985), and Mahajan et al. (1990) studied marketing strategies in new product diffusion models. The production-marketing interface was surveyed by Eliashberg and Steinberg (1993) and Gaimon (1988), while Jørgensen (1982a), Erickson (1995a, 1995b), Moorthy (1993), and Feichtinger et al. (1994) considered their applications to advertising strategies.

The Markov chain approximation method developed by Kushner et al. (1990) and Kushner (2002, 2004) is the current widely used method for a variety of stochastic control problems in continuous time where a Markov chain is generated from a diffusion equation.

Prasad et al. (2003, 2004), Bass, Prasad, and Sethi (2005) have conducted some recent research on stochastic differential game models applied to market competition.

Our survey of the literature indicates the absence of publications that employ a diffusion-jump process to model market share in the market competition area. To our knowledge, the latest report of a stochastic differential game applied to marketing competition is Bass et al. (2005), who studied firms that want to increase the sales of their brands through advertising and that typically have the choice of capturing market share from their competitors through brand advertising, or of increasing primary demand for the category through generic advertising.

1.2.2 Heterogeneity in Markets

In general, heterogeneity is used to refer to the differences among customers. In studying heterogeneity in markets, the use of hierarchical models has been found to be a popular framework. For example, researchers have studied several types of first level models where a regression model is used in the second level. Examples of this approach include a first level normal linear regression model (Blattberg and George (1991)), a first level logit model (Allenby and Lenk (1994), Allenby and Ginter (1995)), a first level probit model (McCulloch and Rossi (1994)), a first level Poisson distribution model (Neelamegham and Chintagunta (1999)), and a first level generalized gamma distribution model (Allenby, Leone, and Jen (1999)). All models show that there is a substantial degree of heterogeneity across units in various marketing data sets. Customer segmentation is also related to heterogeneity. For example, price coefficients could be different for each group of customers. In general, it has been found that wealthy customers are not concerned about price as much as less wealthy customers. In current marketing research literature, no robust regression models are used at the bottom of the hierarchical model. It appears that focusing on level contents may lead to opportunities for improving of results in heterogeneity markets.

Chapter 3 discusses a set of primary factors causing heterogeneity. An instructive list is as follows: (1) brand loyality/pereference heterogeneity, (2) price heterogeneity, (3) structural heterogeneity, (4) demographic heterogeneity, (5) scale heterogeneity, and (6) spatial heterogeneity.

Because too many parameters must be estimated, MCMC is considered as a suitable estimation method (Rowe (2002)). The MCMC method is a simulation technique that generates samples (multiple observations) from the target distribution in the following way: the transition probability of a Markov process is specified by its limiting invariant distribution (target distribution). The Markov chain is then iterated a large number of times, and the outputs are samples from the target distribution. The first method, the Metropolis-Hastings (MH) algorithm is from Metropolis et al. (1953) and Hastings (1970). In this algorithm, the next value of the Markov chain is generated from a proposed density and then accepted or rejected according to the ratio of the density at the candidate point divided by the density at the current point. Another MCMC method is the Gibbs sampling algorithm, introduced by Geman and Geman (1984), in which the next draw is obtained by sampling sub-components of a random vector from a sequence of conditional distributions.

1.2.3 Dynamic Pricing in Production-Inventory Environment

As was shown in Figure (1), an application area for brand choice model considered in this research is dynamic pricing, which is an important method for revenue management (Talluri and van Ryzin (2006)). Some traditional application areas of revenue management include airlines, hotels, car rentals, railroads, etc. Non-traditional application areas are energy, broadcasting, health care, manufacturing, apparel, restaurants, etc.

A firm's revenue management problem can be posed as the maximization of its total expected revenue by selecting appropriate dynamic controls. There are two objectives of this problem: "dynamic pricing" and "capacity control." In the first objective, "dynamic pricing," the company is assumed to be a monopolist with power to influence the demand for each product by varying its price. In this way, the company's policy is to choose a dynamic price for each of its products in order to optimize expected revenue. In the second objective, "capacity control," prices are assumed to be fixed either by the competition or through a higher-order administrator agency. In this case, the company's decision is to choose a dynamic capacity allocation rule that controls when to accept new requests for each of its products. Our work considers three aspects:

(1) The model of a demand process. A demand process is complex, and can not be assumed as a Poisson process with a simple parameter λ, (Gallego and van Ryzin (1994, 1997) and Feng and Gallego (2000)). The demand process is modeled here as a diffusion process.

(2) On a retail level, our model combines the responses of customers. Most models make a common, simplifying, and potentially problematic assumption: that consumer

demand for each product is completely independent of the controls being applied by the seller. Demand processes are modeled by determining which exogenously arriving requests are to be accepted or rejected. Talluri et al. (2004) partially addressed choice behavior issues where demand depends on the current price (the control input). Thus, customers must make a binary choice: to buy or not to buy. This is a logit/probit issue.

(3) On the manufacturing level, the production capacity jump situation is also considered in Chapter 4.

1.3 Contributions of this Work

In this section, we outline our primary contributions after presenting an overview of the research components. The main contributions of this work are: (1) the development of a stochastic differential-jump game (SDJG) model for a more detailed study of brand competition, in which a customer choice behavior is modeled and observed on a strategic management level. The concept of SDJG is developed explicitly and applied to the brand choice problem for the first time. The game model not only models continuous state variables, but also discrete event inputs. (2) A robust hierarchical logit/probit model is proposed for market heterogeneity, which is a more detailed model for the study of choice differences among customers. The robust hierarchical logit/probit model improves significantly log margin density (LMD), which is the classic performance measure in MCMC. Further, this model predicts more precisely compared with a general hierarchical logit/probit model because we have more precise regression estimation. (3) The application of brand choice models in the current topic of dynamic pricing. These models are embedded in production-inventory systems which jump events in production systems are also considered. An explanation of the foregoing contributions is presented in the sequel.

1.3.1 A Stochastic Differential-Jump Game Model

The evolution of market share states can be modeled with a jump-diffusion process when there are discrete event inputs. Brand competition is modeled by a stochastic differential-jump game model in this situation. In a stochastic differential-jump game, state equations are represented as jump-diffusion processes. The specific benefit of this model is that it can model continuous and discontinuous states together. To circumvent the difficulty of obtaining a closed-form solution for the resultant general Hamilton-Jacobi-Bellman (HJB) equation, we introduce the Markov Chain Approximation Algorithm in the examples considered.

The market share of a brand depends on elements such as price, advertising effort, and quality. The objective is to find optimal policies such as advertising effort to maximize total profit. To the best of our knowledge, our model is the first application of a jump-diffusion process in marketing competition research.

Our work begins with the definition of the terminology of a stochastic differential-jump game, followed by the development of a first application of a stochastic differential-jump game model for market competition.

Definition 1.3.1 *Stochastic differential-jump game: the stochastic differential-jump game is a game with state equations including mark-time Poisson processes. The state of dynamic system i at t,*

$$dx_i(t) = f_i(t, x(t), U(t))dt + \sigma_i(t)dW_i(t) + dP_i(t), \quad x(0) = x_0.$$

where $U(t) = (u_1(t), ..., u_{i-1}(t), u_i(t), u_{i+1}(t), ..., u_N(t))$, $x(t) = (x_1(t), ..., x_{i-1}(t), x_i(t), x_{i+1}(t),$ $..., x_N(t))$, $f(x, U, t) = (f_1(x, U, t), ..., f_{i-1}(x, U, t)), f_i(x, U, t), f_{i+1}(x, U, t), .., f_N(x, U, t))$, $W_i(t)$ *is a Brownian motion.* $dP_i(t)$ *is a mark-time Poisson process whose rate is* $\lambda_{J_i}(x, u, t)$ *and whose jump amplitude probability density is* $\eta_i(z)$*. When the system is in state* $(t, x(t))$ *and players select their controls* $u_1(t), ..., u_N(t)$*, player* i *receives the payoff rate* $g_i(x(t), U(t))$*.*

The objective for player i *is*

$$V_i^* = \max_{u_i(x(t))} E\left\{\int_0^T e^{-rt} g_i(x(t), u_i^*(x(t)))dt + e^{-rT}\mathcal{U}(X(T)\right\}$$

where r *is the discount coefficient.*

Definition 1.3.2 *Nash equilibrium: when player* i *expects the other* $N-1$ *players to select their Nash equilibrium strategy. Formally, an N-tuple* $(u_1^*(t), \cdots, u_N^*(t))$ *of strategies constitute a Nash equilibrium if and only if*

$$V_i(u_1^*(t), \cdots, u_i^*(t), \cdots, u_N^*(t)) \geq V_i(u_1^*(t), \cdots, u_{i-1}^*(t), u_i(t), u_{i+1}^*(t), \cdots, u_N^*(t))$$

and

$$\forall u_i^*(t) \in U(t), \quad \forall i \in \{1, ..., N\}$$

We also define

$$u^*(t) = (u_1^*(x(t)), \cdots, u_{i-1}^*(x(t)), u_i^*(x(t)), u_{i+1}^*(x(t)), \cdots, u_N^*(x(t))).$$

Theorem 1.3.1 *Suppose that N-tupe* $(u_1, \cdots, u_N)$ *of functions* $u_i^* : [0, T]$ *is given, and (i) there is a* $x(t) \in X$ *of the initial value problem*

$$dx_i(t) = f_i(x(t), u_i(x(t))dt + \sigma_i(t)dW_i(t) + dP_i(t)$$

(ii) there exists a continuously differentiable function $V_i : [0,T] \times X \longrightarrow R$, such that the following Hamilton-Jacobi-Bellman equations are satisfied for all $(t,x) \in [0,T] \times X$:

$$\frac{\partial V_i^*(x,t)}{\partial t} + H_i(x,t) = rV_i^*(x,t);$$

where

$$H_i(x,t) = \max_{u_i(x(t))} \Big\{ f(x,u_i,t)\frac{\partial V^*(x,t)}{\partial x} + \frac{1}{2}\sum_{j=1}^{n_w}\sum_{i=1}^{n_w} \sigma_{i,j}(t)\frac{\partial^2 V_i^*(x,t)}{\partial x^2}\sigma_{i,j}(t) + g_i(x,u_i) +$$
$$\sum_{k=1}^{n_p} \lambda_{J_k}(x,u_i,t)\int_{\mathcal{Q}} (V^*(x+ez,t) - V^*(x,t))\eta_k(z)dz \Big\}$$
$$u_i^* = u_i^*(x(t)) \tag{1.3.1}$$

where $e = (e_1, e_2, ..., e_N)$ is the $(1 \times N)$ indicating vector, $\sigma_{i,j}(t)$ is the covariance of x_i and x_j.n_w and n_p are the number of Brownian motions and Poisson processes respectively, λ_{J_k} is the rate of kth mark-time Poisson process. $\eta_k(z)$ is the probability density of jump amplitude z of kth mark-time Poisson process. $\mathcal{Q}$ is the domain of z.
(iii) the boundary conditions

$$V_i(T,x) = \mathcal{U}(x_i(T))$$

are satisfied for all $x \in X$ and $i \in \{1,...,N\}$. If $u_i^*(x(t))$ is a maximizer of the right-hand side of the Hamilton-Jacobi-Bellman equation, then $u_i^*(x(t))$ is the Nash equilibrium solution, where σ is the volatility matrix with dimension $(N \times N)$.

The following model is used in the example of Chapter 2.

$$
\begin{aligned}
V_1^*(x) &= \max_{u_1}\left\{ E \int_0^\infty e^{-rt}[m_1 x(t) - c_1 u_1^2(t)]dt \right\} \\
V_2^*(y) &= \max_{u_2}\left\{ E \int_0^\infty e^{-rt}[m_2 y(t) - c_2 u_2^2(t)]dt \right\} \\
dx(t) &= \left(\rho_1\sqrt{1-x(t)} - \rho_2\sqrt{x(t)}\right) dt + \sigma_1(t)dW(t) + dP(t) \\
x(t) + y(t) &= 1 \\
\rho_1 &= \frac{\exp\left(\beta_{10} + u_1(t)\beta_1\right)}{\exp(\beta_{10} + u_1(t)\beta_1) + \exp(\beta_{20} + u_2(t)\beta_2)} \\
\rho_2 &= \frac{\exp\left(\beta_{20} + u_2(t)\beta_2\right)}{\exp(\beta_{10} + u_1(t)\beta_1) + \exp(\beta_{20} + u_2(t)\beta_2)}
\end{aligned}
\tag{1.3.2}
$$

where $x(t)$, $y(t)$ are the sale rates (expressed as a fraction of the total market) at time t, $u(t)$ is the advertising expenditure rate (effort), r is the discount coefficient, and m_1, m_2, c_1, c_2 are profit coefficients and cost coefficients, respectively. ρ_1, ρ_2 are advertisement response coefficients. $\frac{\exp u\beta}{\sum \exp u\beta}$ is a logit model used as a response function, which determines the effectiveness of advertising and pricing effect, while other logit models before $\sqrt{x}$ and $\sqrt{y}$ determine the rate at which consumers are lost due to competitor efforts.

1.3.2 A Robust Hierarchical Bayesian Logit/Probit

Heterogeneity is an important phenomenon in marketing research, but it is difficult to model. A hierarchical structure matches the observation data of heterogeneity. The advantage of this type of model is that it provides a method to scrutinize every detailed element of a problem and heterogeneity model for different customers. A hierarchical model combines the features of products and the characteristics of customers. MCMC is a simulation methodology that makes the Bayesian technique more practical. MCMC has some advantages compared with maximum likelihood estimation (MLE) when the number of parameters is large. MCMC draws samples of those parameters needed to estimate.

Logit and probit models are simple and well-known discrete brand choice models. A hierarchical probit or logit model is a multi-level model in which the top level is the probit or logit, and the bottom level is a regression model. A robust hierarchical Bayesian logit/probit model is presented. By robustness, we mean the model works well with extreme outlier data because of the long tail t distribution used. Robustness makes the estimation of parameters more stable than otherwise possible using MLE.

Our work on this topic includes: (1) a review of heterogeneity in marketing; (2) the development of a robust hierarchical Bayesian logit/probit model; and (3) an experiment to validate our robust model using bank credit card choice data.The experimental results show that robust hierarchical models are better based on LMD and prediction precision. The details of this model are presented in Chapter 3. The model for robust hierarchical logit/probit is

$$\begin{aligned}
&\text{Brand } j \text{ is chosen with probability } p_j, \text{ where}\\
&p_j = logit(x'_{\hat{i}j}\beta_{\hat{i}}) \quad \text{or} \quad p_j = \Phi(x'_{\hat{i}j}\beta_{\hat{i}}),\\
&y_{\hat{i}j} = x'_{\hat{i}j}\beta_{\hat{i}} + \varepsilon_{\hat{i}j},\\
&\beta_{\hat{i}} = \Gamma z_{\hat{i}} + \xi_{\hat{i}},\\
&\varepsilon_{\hat{i}j} \sim Norm(0, \sigma^2),\\
&\xi_{\hat{i}} \sim Norm(0, \Sigma/w_{\hat{i}}),\\
&w_{\hat{i}} \sim \mathcal{G}amma(\nu_1/2, \nu_1/2).
\end{aligned} \tag{1.3.3}$$

and the remaining variables as defined below:

$y_{\hat{i}j}$: the latent variable of profile j valuated by the respondent $\hat{i}$,

$x_{\hat{i}j}$: vector of independent variables associated with profile j for respondent $\hat{i}$, it is a

$(1 \times k)'$ vector, k is the number of properties of a profile,

$\beta_{\hat{i}}$: respondent $\hat{i}$-specific coefficients, $\beta_{\hat{i}} = (\beta_{\hat{i}1}, ..., \beta_{\hat{i}k})'$,

$\varepsilon_{\hat{i}j}$: i.i.d norm error term $Norm(0, \sigma^2)$,

Γ : matrix that relates $\beta_{\hat{i}}$ to the value of $z_{\hat{i}}$,

$z_{\hat{i}}$: respondent's information, is a $(1, 3)$ dimension vector,

$\xi_{\hat{i}}$: unobserved heterogeneity component,

$w_{\hat{i}}$: the weight, Gamma distribution $\mathcal{G}amma(\nu_1/2, \nu_1/2)$.

1.3.3 Dynamic Pricing Models

Dynamic pricing is important to revenue management. Our analysis of dynamic pricing is based on various types of Poisson processes. Because of its complexity and dynamics, the demand process is modeled as a diffusion process. We analyze dynamic pricing problems by considering various scenarios. As an extension of the controlled demand process, certain problems in production and inventory are also investigated. These extended models are solved by employing stochastic dynamic programming.

Work on this topic includes (1) dynamic pricing models presented with λ, and customer choice; (2) an extended dynamic pricing model with a production-inventory system solved using a stochastic optimal control framework; and (3) jump effects on this production-inventory system using a mark-time Poisson process.

In extended production-inventory systems, $\lambda(t)$ is the demand rate, $I(t)$ is the inventory rate, $X(t)$ is the production rate, $h(t)$ is the inventory cost in unit time, $c(t)$ is the production cost in unit time, $q(t)$ is the price, $Q(t)$ is the capacity, ϕ is the demand at each arrival, σ is the volatility, α, θ are constants, and $W(t)$ and $dP(t)$ are Browian motion and mark-time processes respectively. $V^*(X, \lambda)$ is the maximum cumulative profit from 0 to t T at state X, λ. Then, our resultant model may be represented as:

$$
\begin{aligned}
V^*(X,\lambda) &= \max_q E\int_0^T \left\{ q(t)\min[\lambda(t)logit(q(t)\beta)\phi, X(t)] - h(t)I(t) - c(t)X(t) \right\} dt \\
&\text{s.t.} \\
dX(t) &= f(X(t),\lambda(t),q(t))dt + dP(t) \\
d\lambda(t) &= \alpha\left(\theta - \lambda(t)\right)dt + \sigma\sqrt{\lambda(t)}dW(t) \\
I(t) &= \max[X(t) - \lambda(t)logit(q(t))\phi, 0] \\
I_0 &= 0, X(t) \le Q(t), I(t), X(t), \lambda(t) \in R^+.
\end{aligned}
\tag{1.3.4}
$$

or

$$
\begin{aligned}
V^*(X,\lambda) &= \max_q E\int_0^T \left\{ q(t)\min[\lambda(t)probit(q(t)\beta)\phi, X(t)] - h(t)I(t) - c(t)X(t) \right\} dt \\
&\text{s.t.} \\
dX(t) &= f(X(t),\lambda(t),q(t))dt + dP(t) \\
d\lambda(t) &= \alpha\left(\theta - \lambda(t)\right)dt + \sigma\sqrt{\lambda(t)}dW(t) \\
I(t) &= \max[X(t) - \lambda(t)probit(q(t))\phi, 0] \\
I_0 &= 0, X(t) \le Q(t), I(t), X(t), \lambda(t) \in R^+.
\end{aligned}
\tag{1.3.5}
$$

Conclusion: In this dissertation, we develop a new game model, namely, stochastic differential-jump game model and apply it to the brand's market competition in a specific situation. Our robust hierarchical Bayesian logit/probit model has more precise estimation and predictive ability. In our examples in Chapter 3, the AIC, BIC, and LMD are improved by 60% in robust regression level. The prediction error is reduced by around 80%. Our continuous-time dynamic pricing model in Chapter 4 is suitable in different situations: jump, or no-jump cases.

CHAPTER II

A STOCHASTIC DIFFERENTIAL-JUMP GAME MODEL FOR BRAND COMPETITION

A brand's market share depends on elements such as price, advertising, quality, entry of new brands into the market, and availability of funds. The introduction of new brands and the availability of funds can be viewed as discrete events where market share states can be modeled by a jump-diffusion process.

The objective of this study is to find optimal policies such as pricing or advertising effort in some jump event situations using a stochastic differential-jump game model where state equations are jump-diffusion processes. The specific benefit of this model is the ability to model continuous states and discrete events together. An extensive review of the literature indicates that this is the first application of a jump-diffusion process in market competition.

In the current research, brand competition is modeled using a stochastic differential-jump game model. Due to the difficulty of using a closed-form solution for a general Hamilton-Jacobi-Bellman (HJB) equation, the Markov chain approximation algorithm is used in these examples.

This chapter considers (1) a definition of terminology and a framework for a stochastic differential-jump game model and (2) the first application of a stochastic differential-jump game model for marketing competition.

2.1 Introduction

Von Neumann and Morgenstern laid the foundation of modern game theory in their book "*Theory of Games and Economics Behavior*" published in 1944. Since then, the application of game theoretic analysis in many fields such as economics, biology, engineering, operations research, and politics has become more popular.

Dynamic game theory, including differential games and stochastic differential games, has been developed from static games. In this chapter, we present explicitly the definition of a stochastic differential-jump game.

Differential games belong to a subset of the general class of dynamic games originating from systems theory, optimal control, and dynamic programming. Such games include a number of state variables that describe states of a dynamic system over a period of time. Historically, differential game theories were developed by Isaacs, Friedman, Leitmann, Krasovskii, Subbotin, Basar, and Olsder. Differential games are popularly used in economic and management applications. Jørgensen (1982a, 1982b, 1982c) summarized some applications of differential games in advertising. In the early 1980s, only a few works dealt with pricing, but they led to continued research of competitive advertising and pricing strategies, for example, the application of games theory in marketing channels, as well as the interaction of marketing with other functional areas such as production, capacity expansion, and finance. Moorthy (1985, 1993) and Rao (1990) surveyed game theoretic modelings in marketing. Jørgensen (1986), Rao (1990), and Moorthy (1993) discussed pricing strategies. Dolan et al. (1985), and Mahajan et al. (1990) studied marketing strategies in new product diffusion models. The production-marketing interface is surveyed by Eliashberg and Steinberg (1993), and Gaimon (1988). Jørgensen (1982a), Erickson (1995a, 1995b), Moorthy (1993), and Feichtinger et al. (1994) studied advertising strategies.

A stochastic differential game is a differential game plus a stochastic item (Brownian motion) in its state equation. Cellini and Lambertini (2003) illustrated a differential oligopoly game where firms compete with homogeneous goods in the market phase and invest in advertising activities aimed at increasing the consumers' reservation price. They derived the open-loop and closed-loop Nash equilibrium, and showed that the properties of the equilibrium depend on the curvature of the market demand function. The comparative assessment of these equilibria showed that the firms' advertising efforts are greater in the open-loop equilibrium than in the closed-loop equilibrium.

Prasad and Sethi (2004) analyzed optimal advertising decisions in a duopolistic market, where each firm's market share was assumed to depend on its own and its competitor's advertising decisions, and was subject to stochastic disturbances. Their model for advertising was based on Sethi's stochastic advertising model and the Lanchester model of combat. Bass et al. (2005) studied a similar model for firms that wanted to increase the sales of their brands through advertising.

Pricing and advertising policies in market situations may be exposed to external jump events. External jump events exist in the real world. For example, hurricanes create an emergency need for many products. In this case, we can reduce the effort of advertising. If there are new entrants in the market, advertising efforts need to be increased to maintain market share. In addition to those random events, some planned events can also affect the system. The occurrence of this type of event is modeled by the mark-time Poisson process.

A stochastic differential-jump game model is a game model where state processes are defined by a jump-diffusion process. Jump-diffusion processes were first used by Merton in optimal investment and more recently were used in asset pricing as well. But a jump-diffusion process is seldom used in the game framework to study price policy, and advertising effort in external jump event situations. To the best of our knowledge, the current research is the first time the stochastic differential-jump game model has been developed and applied in market competition.

2.2 Review of Market Share Model

In this section, the advertising effect and market share model is reviewed. Consider an N-firm oligopoly and let $x_i(t)$ denote the market share of firm $i \in \{1, ..., N\}$.

The state space $\tilde{X}$ is represented by

$$\tilde{X} = \{x_i(t) \in R | x_i(t) \in [0, 1], i \in \{1, ..., N\}, \sum_{i=1}^{N} x_i(t) = 1\}$$

Although Lanchester (1956) developed a model for military combat, the model has also been found useful in marketing competition. The advertising effort rate of firm i at time t is denoted by $a_i(t)$. Let the initial market shares be $x_i(0) = x_{i0}$ such that $(x_{10}, ..., x_{N0}) \in \tilde{X}$. $f_i(a_i(t))$ is the market share gain by the advertising effort of competitor i.

A general version of the Lanchester dynamics is

$$\frac{dx_i(t)}{dt} = [1 - x_i(t)]f_i(a_i(t)) - x_i(t) \sum_{j=1, j\neq i}^{N} f_j(a_j(t)) = f_i(a_i(t)) - x_i(t) \sum_{j=1}^{N} f_j(a_j(t))$$

If the entire industry sales are fixed and equal to m in a market, we have

$$\sum_{i=1}^{N} x_i(t) = m$$

$$\frac{dx_i(t)}{dt} = [m - x_i(t)]f_i(a_i(t)) - x_i(t) \sum_{j=1, j \neq i}^{N} f_j(a_j(t)) = mf_i(a_i(t)) - x_i(t) \sum_{j=1}^{N} f_j(a_j(t))$$

The objective function of the differential game model for the Lanchester dynamics is

$$V_i = \int_0^T e^{-r_i t}[\pi_i x_i(t) - C_i(a_i(t))]dt + e^{-r_i T}\mathcal{U}_i(x_i(T))$$

π_i is the constant unit margin of firm i and $C_i(a_i(t))$ is the cost of advertising effort. $a_i(t)$ is the policy of firm i. $\mathcal{U}_i(x_i(T))$ is the utility function of the final state. The policy could be the price or advertising effort.

The Vidale-Wolfe model (Vidale and Wolfe (1957)) is

$$\frac{dx(t)}{dt} = \rho u(t)[m - s(t)] - \delta x(t)$$

where $u(t)$ is the advertising expenditure rate, ρ is a response constant, and δ is a market share decay constant.

The original Sethi model (Sethi (1983)) is stochastic.

$$dx(t) = \left[\rho u(t)\sqrt{1 - x(t)} - \delta x(t)\right] dt + \sigma(x)dW(t)$$

where $x(t)$ is the sales rate (expressed as a fraction of the total market), $u(t)$ is the advertising expenditure rate, ρ is a response constant, and δ is a market share decay constant. The parameter ρ may be conceptualized as brand characteristics, determining the result of advertising, while δ determines the rate at which consumers are lost as the result of a competitor's effort. $W(t)$ is the Brownian motion.

Sorger (1989) combined the Sethi model with the Lanchester framework to obtain

$$\frac{dx(t)}{dt} = \rho_1 u(x,y)\sqrt{1-x(t)} - \rho_2 u(x,y)\sqrt{x(t)}$$
$$\frac{dy(t)}{dt} = \rho_1 u(x,y)\sqrt{1-y(t)} - \rho_2 u(x,y)\sqrt{y(t)}$$

where $x(t), y(t)$ are the sales rate (expressed as a fraction of the total market); $u(x,y)$, abbreviations for $u(x(t), y(t))$ is the advertising expenditure rate, and ρ_1, ρ_2 are response parameters.

Prasad and Sethi (2003) considered a duopoly extension of the Sethi model. The deterministic part of their dynamics differs from Sorger's model in having a parameter δ,a market share decay constant, and is given by

$$\frac{dx(t)}{dt} = \rho_1 u(x,y)\sqrt{1-x(t)} - \rho_2 u(x,y)\sqrt{x(t)} - \delta\Big(x(t) - y(t)\Big)$$
$$\frac{dy(t)}{dt} = \rho_1 u(x,y)\sqrt{1-y(t)} - \rho_2 u(x,y)\sqrt{y(t)} - \delta\Big(y(t) - x(t)\Big)$$

Prasad and Sethi (2004) considered the following model, including the stochastic noise.

$$dx(t) = \left[\rho_1 u_1(x,y)\sqrt{1-x(t)} - \rho_2 u_2(x,y)\sqrt{x(t)} - \delta\Big(x(t) - y(t)\Big)\right] dt + \sigma(x,y) dW(t)$$
$$dy(t) = \left[\rho_2 u_2(x,y)\sqrt{1-y(t)} - \rho_1 u_1(x,y)\sqrt{y(t)} - \delta\Big(y(t) - x(t)\Big)\right] dt - \sigma(x,y) dW(t)$$

The meaning of these parameters follows the definitions of the Sethi model. ρ_1, ρ_2 are response parameters. δ is a market share decay constant. $\sigma(x,y)$ is a volatility.

2.3 Markov Chain Approximation Algorithm

The finite difference method is a basic method to solve partial differential equations. The Markov chain approximation method was invented by Kushner and Dupuis (1990) and Kushner (2002, 2004). The Markov chain approximation method comes from the finite difference method. The primary characteristic of the Markov chain approximation method is its nice convergence property.

The system is defined by $dx(t) = f(x,u)dt + \sigma(x)dW(t)$. The objective function is:

$$V^*(x_0, t_0) = \max\left\{V(x_0, t_0) = E\int_{t_0}^{t_f} g(x(t), u(t), t)dt\right\}$$

where $x(t)$ is the state variable, $u(t)$ is a control variable, and $W(t)$ is a Brownian motion process. The Hamilton-Jacob-Bellman equation of the system is

$$0 = \frac{\partial V^*(x,t)}{\partial t} + \max_u \left\{ f(x,u)\frac{\partial V^*(x,t)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 V^*(x,t)}{\partial x^2} + g(x,u,t)\right\} \tag{2.3.1}$$

Here, the derivative can be calculated by the finite difference. The central difference for the second derivative is

$$\frac{\partial^2 V^*(x,t)}{\partial x^2} = \frac{V^*(x+h,t) - 2V^*(x,t) + V^*(x-h,t)}{h^2}$$

Equation (2.3.1) can be approximated by the backward Euler approximation. h is the step length in the x direction.

$$V^*(x, t_{k-1}) = V^*(x, t_k) + \Delta t_k \max_u\left\{ f_k(x,u)\frac{\partial V^*}{\partial t}(x,t_k) + \frac{1}{2}\sigma^2(x)\frac{\partial^2 V^*}{\partial x^2}(x,t_k) + g(x,u,t_k)\right\} \tag{2.3.2}$$

We use the finite difference for first order and second derivatives on equation (2.3.2). So,

$$\begin{aligned} V^*(x, t_{k-1}) = \max_u\Big\{ & p_k(x, x|u_{k-1})V^*(x,t_k) + p_k(x, x+h|u_{k-1})V^*(x+h, t_k) \\ & + p_k(x, x-h|u_{k-1})V^*(x-h,t_k) + \Delta t_k g(x,u,t_k)\Big\} \end{aligned}$$

where these transition probabilities are

$$\begin{aligned}
p_k(x,x|u_{k-1}) &= 1-\Delta t_k(\frac{\sigma_k^2}{h^2}+\frac{|f_k|}{h}) \\
p_k(x,x-h|u_{k-1}) &= \Delta t_k(\frac{\sigma_k^2}{2h^2}+\frac{[f_k]_+}{h}), \quad [f_k]_+ = \max\{+f_k,0\} \\
p_k(x,x+h|u_{k-1}) &= \Delta t_k(\frac{\sigma_k^2}{2h^2}+\frac{[f_k]_-}{h}), \quad [f_k]_- = \max\{-f_k,0\} \\
\Delta t_k &\le \frac{h^2}{\sigma_k^2+h|f_k|}, \quad f_k = f(t_k,u)
\end{aligned} \tag{2.3.3}$$

Equation (2.3.3) is the exact result in Hanson (2006). Kushner and Dupuis (2001), provided some conditions for the algorithm's convergence.

2.4 Introduction to a Stochastic Differential-Jump Game

A jump event is added to the stochastic differential game model so that it becomes a stochastic differential-jump game (SDJG). Jump-diffusion processes are widely used in mathematical finance, especially in credit risk analysis.

This model is used for brand competition when there are uncertain events in markets, scheduled or random, and impulse inputs with a positive or negative impact. This model has more capability than a stochastic differential game model because it uses only diffusion processes. A diffusion process is a continuous process that can not model the discrete change. The mark-time Poisson process makes up the shortage of a diffusion process.

Definition 2.4.1 *Stochastic differential-jump game: A stochastic differential-jump game can be defined simply as a stochastic differential game with jump events. The state of a dynamic system at t,*

$$dx_i(t) = f_i(x,U,t)dt + \sigma_i(t)dW_i(t) + dP_i(t), \quad x_i(0) = x_0.$$

where $U(t) = (u_1(t), ..., u_{i-1}(t), u_i(t), u_{i+1}(t), ..., u_N(t))$, $x(t) = (x_1(t), ..., x_{i-1}(t), x_i(t),$ $x_{i+1}(t), ..., x_N(t))$, $f(x, U, t) = (f_1(x, U, t), ..., f_{i-1}(x, U, t)), f_i(x, U, t), f_{i+1}(x, U, t)$ $, ..., f_N(x, U, t))$, $W_i(t)$ *is a Brownian motion.* $dP_i(t)$ *is a mark-time Poisson process whose rate is* $\lambda_{J_i}(x, u, t)$ *and whose jump amplitude probability density is* $\eta_i(z)$*. When the system is in state* $(t, x(t))$ *and the players select their controls* $u_1(t), ..., u_N(t)$*, player* i *receives the payoff rate* $g_i(x, U, t)$*.*

The objective function is,

$$V_i^*(x,t) = \max_{u_i} V_i(x,t) = \max_{u_i} E\left\{e^{-r(T-t)}\mathcal{U}_i(X(T)) + \int_t^T e^{-r(\tau-t)} g_i(x(\tau), u_i^*(\tau), \tau) d\tau\right\}$$

where r is the discount coefficient.

Definition 2.4.2 *Nash Eequilibrium: when player* i *expects the other* $N-1$ *players to select their Nash equilibrium strategy. Formally, an N-tuple* $(u_1^*(t), \cdots, u_N^*(t))$ *of strategies constitutes a Nash equilibrium if and only if*

$$V_i(u_1^*(t), \cdots, u_i^*(t), \cdots, u_N^*(t)) \geq V_i(u_1^*(t), \cdots, u_{i-1}^*(t), u_i(t), u_{i+1}^*(t), \cdots, u_N^*(t))$$

We also define

$$u^*(t) = (u_1^*(x(t)), \cdots, u_{i-1}^*(x(t)), u_i^*(x(t)), u_{i+1}^*(x(t)), \cdots, u_N^*(x(t))).$$

Theorem 2.4.1 *Suppose that N-tupe* $(u_1, \cdots, u_N)$ *of functions* $u_i^* : [0, T]$ *is given, and (i) there is a* $x(t) \in X$ *of the initial value problem*

$$dx_i(t) = f_i(x(t), u_i(x(t))dt + \sigma_i(t)dW_i(t) + dP_i(t)$$

(ii) there exists a continuously differentiable function $V_i : [0, T] \times X \longrightarrow R$*, such that the following Hamilton-Jacobi-Bellman equations are satisfied for all* $(t, x) \in [0, T] \times X$*:*

$$\frac{\partial V_i^*(x,t)}{\partial t} + H_i(x,t) = rV_i^*(x,t);$$

where

$$H_i(x,t) = \max_{u_i(x(t))}\Bigg\{f(x,u_i,t)\frac{\partial V^*(x,t)}{\partial x} + \frac{1}{2}\sum_{j=1}^{n_w}\sum_{i=1}^{n_w}\sigma_{i,j}(t)\frac{\partial^2 V_i^*(x,t)}{\partial x^2}\sigma_{i,j}(t) + g_i(x,u_i)+$$
$$\sum_{k=1}^{n_p}\lambda_{J_k}(x,u_i,t)\int_{\mathcal{Q}}(V^*(x+ez,t)-V^*(x,t))\eta_k(z)dz\Bigg\}$$
$$u_i^* = u_i^*(x(t)) \tag{2.4.1}$$

and where $e = (e_1, e_2, ..., e_N)$ *is the* $(1 \times N)$ *indicating vector,* $\sigma_{i,j}(t)$ *is the covariance of* x_i *and* x_j*, and furthermore* n_w *and* n_p *are the number of Brownian motions and Poisson processes, respectively,* λ_{J_k} *is the rate of the kth mark-time Poisson process,* $\eta_k(z)$ *is the probability density of jump amplitude* z *of the kth mark-time Poisson process, and* $\mathcal{Q}$ *is the domain of* z*.*

(iii) the boundary conditions

$$V_i(T,x) = \mathcal{U}(x_i(T)))$$

are satisfied for all $x \in X$ *and* $i \in \{1,...,N\}$*.*

If $u_i^*(x(t))$ *is a maximizer of the right-hand side of the Hamilton-Jacobi-Bellman equation, then* $u_i^*(x(t))$ *is the Nash equilibrium solution, where* σ *is the volatility matrix with dimension* $(N \times N)$*.*

Proof. If there are two solutions $u_i(t)$ and $u_i^*(t)$(optimal solution), and their trajectories are $x(t)$ and $x^*(t)$, respectively.

$$u(t) = (u_1^*(x(t)),\cdots,u_{i-1}^*(x(t)),u_i(x(t)),u_{i+1}^*(x(t)),\cdots,u_N^*(x(t)))$$

The $u_i^*(t)$ is the maximizer because it is assumed as the optimal solution.

$$\frac{\partial V_i(x,t)}{\partial t} + \tilde{H}_i(x,t) \le rV_i(x,t) \tag{2.4.2}$$

where

$$\tilde{H}_i(x,t) = \left\{ f(x,u_i,t)\frac{\partial V}{\partial x} + \frac{1}{2}\sum_{j=1}^{n_w}\sum_{i=1}^{n_w}\sigma_{i,j}(t)\frac{\partial^2 V_i}{\partial x^2}(x,t)\sigma_{i,j}(t) + g_i(x,u_i) + \right.$$
$$\left. \sum_{k=1}^{n_p}\lambda_{J_k}(x,u_i,t)\int_{\mathcal{Q}}(V(x+ez,t)-V(x,t))\eta_k(z)dz \right\}$$

and

$$\frac{\partial V_i^*(x,t)}{\partial t} + H_i(x,t) = rV_i^*(x,t) \tag{2.4.3}$$

from equation (2.4.2) and (2.4.3),

$$\int_0^T g_i(x,u_i,t)dt + V_i(x(T),T) - V_i(x(0),0) \leq \int_0^T rV_i(x(t),t)dt$$

$$\int_0^T g_i(x^*,u_i^*,t)dt + V_i^*(x^*(T),T) - V_i^*(x(0),0) = \int_0^T rV_i^*(x^*(t),t)dt$$

So

$$\int_0^T g_i(x^*,u_i^*,t)dt + V_i^*(x^*(T),T) \geq \int_0^T g_i(x,u_i,t)dt + V_i(x(T),T)$$

The following is to validate the H_i by the principle of optimality.

$$V_i^*(x,t) = \max_{u_i} E\left\{ g_i(x,u,t)dt + \frac{1}{1+rdt}V_i^*(x,t+dt)\middle| x(t)=x, u(t)=u \right\}$$
$$(1+rdt)V_i^*(x,t) = \max_{u_i} E\left\{ g_i(x,u,t)(1+rdt)dt + V_i^*(x,t+dt)\middle| x(t)=x, u(t)=u \right\}$$
$$(1+rdt)V_i^*(x,t) = \max_{u_i} E\left\{ g_i(x,u,t)(1+rdt)dt + V_i^*(x,t) + \frac{\partial V_i^*(x,t)}{\partial t}dt \right.$$
$$+ \frac{\partial V_i^*(x,t)}{\partial x}\Big(f(x,u,t)dt + \sigma_i(t)dW_i(t)\Big)$$
$$\left. + \frac{1}{2}\sigma^2\frac{\partial^2 V_i^*(x,t)}{\partial x^2}dt + \sum_k \int_{\mathcal{Q}}\Big[V_i^*(x+ez,t) - V_i^*(x,t)\Big]\mathcal{P}_k \right\}$$

$$
\begin{aligned}
(1+rdt)V_i^*(x,t) &= \max_{u_i}\bigg\{ g_i(x,u,t)(1+rdt)dt + V_i^*(x,t) + \frac{\partial V_i^*(x,t)}{\partial t}dt + \frac{\partial V_i^*(x,t)}{\partial x}f(x,u,t)dt \\
&+ \frac{1}{2}\sigma^2\frac{\partial^2 V_i^*}{\partial x^2}(x,t)dt + \sum_k \int_{\mathcal{Q}}\Big[V_i^*(x+ez,t) - V_i^*(x,t)\Big]\eta_k(z)dz\lambda_{J_k}(x,u,t)dt\bigg\} \\
rV_i^*(x,t) &= \max_{u_i}\bigg\{ g_i(x,u,t)(1+rdt) + \frac{1}{dt}\bigg(\frac{\partial V_i^*(x,t)}{\partial t}dt \\
&+ \frac{\partial V_i^*(x,t)}{\partial x}f(x,u,t)dt + \frac{1}{2}\sigma^2\frac{\partial^2 V_i^*(x,t)}{\partial x^2}dt \\
&+ \sum_k \int_{\mathcal{Q}}\Big[V_i^*(x+ez,t) - V_i^*(x,t)\Big]\eta_k(z)dz\lambda_{J_k}(x,u,t)dt\bigg)\bigg\} \\
&\text{because of} \quad dt \to 0 \\
&= \max_{u_i}\bigg\{ g_i(x,u,t) + \frac{\partial V_i^*(x,t)}{\partial t} + \frac{\partial V_i^*(x,t)}{\partial x}f(x,u,t) + \frac{1}{2}\sigma^2\frac{\partial^2 V_i^*(x,t)}{\partial x^2} \\
&+ \sum_k \int_{\mathcal{Q}}\lambda_{J_k}(x,u,t)\Big[V_i^*(x+ez,t) - V_i^*(x,t)\Big]\eta_k(z)dz\bigg\} \\
rV_i^*(x,t) &= \max_{u_i}\bigg\{ \frac{\partial V_i^*(x,t)}{\partial t} + g_i(x,u,t) + \frac{\partial V_i^*}{\partial x}(x,t)f_i(x,u,t) + \frac{1}{2}\sigma^2\frac{\partial^2 V_i^*(x,t)}{\partial x^2} \\
&+ \sum_k \lambda_{J_k}(x,u,t)\int_{\mathcal{Q}}\Big[V_i^*(x+ez,t) - V_i^*(x,t)\Big]\eta_k(z)dz\bigg\} \\
u &= \{u_1^*, ..., u_{i-1}^*, u_i, u_{i+1}^*, ..., u_N^*\}
\end{aligned}
\tag{2.4.4}
$$

□

The equation above can be solved using the Markov chain approximation method when $n_p = 1$. $h = (h_1, h_2, ..., h_N)$ is the step length in the x direction. $e = (e_1, e_2, ..., e_N)$ is the $(1 \times N)$ indicating vector. $f_{jk} = f_j(x, U, t_k)$, $\sigma_{jk} = \sigma_j(t_k)$.

$$
\begin{aligned}
rV_i^*(x,t) &= \max_{u_i}\bigg\{ \frac{\partial V_i^*}{\partial t}(x,t) + f(x,u_i^*,t)\frac{\partial V_i^*}{\partial x}(x,t) + \frac{1}{2}\sigma^2(t)\frac{\partial^2 V_i^*}{\partial x^2}(x,t) + \\
&\sum_j \int_{\mathcal{Q}}\lambda_{J_j}(x,u_i^*,t)\Big[V_i^*(x+ez,t) - V_i^*(x,t)\Big]\eta(z)dz + g_i(x,u_i^*,t)\bigg\} \\
V_i^*(x,t_{k-1}) &= \max_{u_i}\bigg\{ \Delta t_k g_i(x,u_i^*,t_{k-1}) + \sum \lambda_{J_j}(x,u_i^*,t_k)\Delta t_k \mathcal{J}_j + p_k(x,x|u_i^*)V_i^*(x,t_k) + \\
&\sum_j \Big(p_k(x_j, x_j+h_j|u_i^*)V_i^*(x+e_jh,t_k) + p_k(x_j,x_j-h_j|u_i^*)V_i^*(x-e_jh,t_k)\Big)\bigg\}
\end{aligned}
$$

Transition probability at time k from state x_i, to x_i, $x_i - h_i$, and $x_i + h_i$ are

$$
\begin{aligned}
p_k(x_i, x_i|u_i^*) &= 1 - \Delta t_k r - \Delta t_k \sum_i (\frac{\sigma_{ik}^2}{h_i^2} + \frac{|f_{jk}|}{h_i}) \\
p_k(x_i, x_i - h_i|u_i^*) &= \Delta t_k (\frac{\sigma_{ik}^2}{2h_i^2} + \frac{[f_{ik}]_+}{h_j}), \quad [f_{ik}]_+ = \max\{+f_{ik}, 0\} \\
p_k(x_i, x_i + h_i|u_i^*) &= \Delta t_k (\frac{\sigma_{ik}^2}{2h_i^2} + \frac{[f_{ik}]_-}{h_i}), \quad [f_{ik}]_- = \max\{-f_{ik}, 0\} \\
\mathcal{J}_j &= \int_{\mathcal{Q}} \Big[V_i^*(x + ez, t_k) - V_i^*(x(t)) \Big] \eta_j(z) dz \\
\Delta t_k &\leq \frac{h^2}{\sum_j (\sigma_{ik}^2 + h_j |f_{ik}|)}
\end{aligned}
$$

The above model is a Markov decision process model. The state variable is x_i. u_i is a decision variable, k is a stage variable (time), $V_i^*(x, t_{k-1})$ is the optimal return function of player i at stage k.

From the Markov decision process framework, the complexity of the method is $\mathcal{O}(\tilde{N}\tilde{D}\tilde{T})$ from the Markov property, where $\tilde{T}$ is the number of the stage. $\tilde{N}$ is the number of states in each stage and $\tilde{D}$ is the number of decisions in each stage.

A Markov decision process has the contraction mapping property in terms of a value function. For example, if there are two value functions that will go thorough the Markov decision process one time, then we have

$$||\tilde{\Gamma}W - \tilde{\Gamma}V|| \leq \tilde{\alpha}||W - V||$$

and the $||W|| = max_S|W(S)|$, $0 \leq \tilde{\alpha} \leq 1$, $\tilde{\Gamma}$ is the value iteration operator of a MDP. Moreover, the difference of the two value functions after the nth iterations in the two functions satisfies the bound

$$||V_n - V|| \leq \frac{\tilde{\alpha}^n ||V_1 - V_0||}{1 - \tilde{\alpha}} = \tilde{\epsilon}$$

So, the total calculations giving $\tilde{\epsilon}$ is $\mathcal{O}(NDln(\tilde{\epsilon} * (1-\tilde{\alpha})/(|V_1 - V_0| * \tilde{\alpha}))$.

When $T \to \infty$, the above model becomes an infinite horizon model. We use the infinite horizon model in all examples in this chapter.

Theorem 2.4.2 *In the infinite horizon case, the objective function is*

$$V_i^*(x) = \max_{u_i} E\left\{\int_0^\infty e^{-r(\tau)} g_i(x(\tau), u_i^*(\tau), \tau) d\tau\right\} \tag{2.4.5}$$

where E is the mean about x. We have the HJB:

$$H_i(x) = rV_i^*(x)$$

and

$$H_i(x) = \max_{u_i(x)} \Bigg\{ f(x, u_i) \frac{\partial V^*}{\partial x} + \frac{1}{2} \sum_{j=1}^{n_w} \sum_{i=1}^{n_w} \sigma_{i,j} \frac{\partial^2 V_i^*}{\partial x^2}(x) \sigma_{i,j} + g_i(x, u_i) + \tag{2.4.6}$$

$$\sum_{k=1}^{n_p} \lambda_{J_k}(x, u_i, t) \int_{\mathcal{Q}} (V_i^*(x + ez) - V_i^*(x)) \eta_k(z) dz \Bigg\}$$

Proof.

$$
\begin{aligned}
V_i(x) &= \max_{u_i(x)} \left\{ E \int_0^{\Delta t} e^{-rt} g_i(x(\tau), u_i^*(\tau), \tau) d\tau + E \int_{\Delta t}^{\infty} e^{-rt} g_i(x(\tau), u_i^*(\tau), \tau) d\tau \right\} \\
&= \max_{u_i(x)} \left\{ E \int_0^{\Delta t} e^{-rt} g_i(x(\tau), u_i^*(\tau), \tau) d\tau + E e^{-r\Delta t} \int_{\Delta t}^{\infty} e^{-r(\tau - \Delta t)} g_i(x(\tau), u_i^*(\tau), \tau) d\tau \right\} \\
&= \max_{u_i(x)} \left\{ E \int_0^{\Delta t} e^{-rt} g_i(x(\tau), u_i^*(\tau), \tau) d\tau + E e^{-r\Delta t} V_i(x(\Delta t)) \right\} \\
0 &= \frac{1}{\Delta t} \max_{u_i(x)} \left\{ E e^{-r\Delta t} V_i(x(\Delta t)) - V_i(x) + E \int_0^{\Delta t} e^{-rt} g_i(x(\tau), u_i^*(\tau), \tau) d\tau \right\} \\
0 &= E \frac{V_i(x(\Delta t)) - V_i(x)}{\Delta t} + \frac{e^{-r\Delta t} - 1}{\Delta t} E V_i(x(\Delta t)) + E \int_0^{\Delta t} e^{-rt} g_i(x(\tau), u_i^*(\tau), \tau) d\tau
\end{aligned}
$$

by the It$\widehat{o}$ lemma with jump when $\Delta t \longrightarrow 0$

$$
H_i(x) = rV_i^*(x)
$$

$$
H_i(x) = \max_{u_i(x)} \left\{ g_i(x, u_i) + f(x, u_i) \frac{\partial V^*}{\partial x} + \frac{1}{2} \sum_{j=1}^{n_w} \sum_{i=1}^{n_w} \sigma_{i,j} \frac{\partial^2 V_i^*}{\partial x^2}(x) \sigma_{i,j} + \right.
$$

$$
\left. \sum_{k=1}^{n_p} \lambda_{J_k}(x, u_i, t) \int_{\mathcal{Q}} (V^*(x + ez) - V^*(x)) \eta_k(z) dz \right\}
$$

□

We use the results in our examples in this chapter. We develop three examples about stochastic optimal control market share with jump events and without jump events. All examples are infinite horizon cases.

2.4.1 Properties of Nash Equilibrium Optimal Solution

We discuss the properties of the Nash equilibrium solution. The system is:

$$
\begin{aligned}
&\max_u V(u) = E \int_0^{t_f} G(x, u, t) dt \\
&\qquad \text{s.t.} \\
&\qquad dx(t) = A(t)dt + \sigma(t)dW(t) + dP(t) \\
&\qquad x(t_f) = x(t_0) + \int_0^{t_f} A(t)dt + \int_0^{t_f} \sigma(t)dW(t) + \int_0^{t_f} dP(t)
\end{aligned}
\tag{2.4.7}
$$

The model definition follows that of Esogbue et al. (2006). $V(u) = (V_1(u), V_2(u), ..., V_N(u))$ is a vector with dimension m, $G(x, u, t) = (g_1(x, u, t), g_2(x, u, t), ..., g_N(x, u, t))$ for

$N \geq 2$, and $G_i(.,.,.)$ are assumed to be continuously differentiable concave functions of x and u. $x(t) = (x_1(t), ...x_{i-1}(t), x_i(t), x_{i+1}(t), ..., x_m(t))$, $dW(t) = (dW_1(t), dW_2(t), ..., dW_N(t))$, $dP(t) = (dP_1(t), dP_2(t), ..., dP_N(t))$. $dP_i(t)$ is the mark-time Poisson process with the measure $E[\mathcal{P}_i(dt, dz)] = \lambda_{J_i} dt\eta(z)dz$,

Theorem 2.4.3 *Let $u^*(t)$ be an admissible solution (control): then $u^*(t)$ is not a Nash equilibrium optimal solution if there exists a piecewise continuous function $du(t) \in R^N$ such that the following inequality holds:*

$$\left[\int_0^{t_f} \frac{\partial G}{\partial x} A(t)dt + \int_0^{t_f} \lambda_J(x(t)) \int_{\mathcal{Q}} \Big(G(x+z,u,t) - G(x,u,t)\Big)\eta(z)dzdt + \int_0^{t_f} \frac{\partial G}{\partial u} du(t)\right] > 0 \tag{2.4.8}$$

Proof.

Sufficiency: Suppose there exists a piecewise continuous function $du(t) \in R^N$ such that the functional inequality holds. Let $u(t) = u^*(t) + du(t)$. Then,

$$V(u) - V(u^*) = E\int_0^t [G(x,u,t) - G(x^*,u^*,t)]dt \tag{2.4.9}$$

where $x^*(t)$ is the state vector of system (2.4.7) when $u(t) = u^*(t)$. Expanding the integrand in (2.4.9), $\forall t \in [0, t_f]$ with the It$\widehat{o}$ lemma , then leads to the following:

$$\begin{aligned}
\triangle V &= G(x,u,t) - G(x^*,u^*,t) \\
&= G\Big(x(t_0) + \int_0^t A(t)dt + \int_0^t \sigma(t)dW(t) + \int_0^t dP(t), u, t\Big) - G(x^*,u^*,t) \\
&= \frac{\partial G}{\partial x}\Big(A(t)dt + \sigma(t)dW(t)\Big) + \frac{\partial G}{\partial u}du(t) + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\sigma^2(t)dt \\
&+ \int_{\mathcal{Q}}\Big(G(x+z,u,t) - G(x,u,t)\Big)\mathcal{P}(t)
\end{aligned} \tag{2.4.10}$$

$$
\begin{aligned}
V(u)-V(u^*) &= E\int_0^{t_f}[G(x,u,t)-G(x^*,u^*,t)]dt \\
&= E\Big[\int_0^{t_f}\frac{\partial G}{\partial x}\Big(A(t)dt+\sigma(t)dW(t)\Big)+\int_0^{t_f}\frac{1}{2}\frac{\partial G}{\partial x^2}\sigma^2(t)dt \\
&+\int_0^{t_f}\lambda_J(x(t))\int_{\mathcal{Q}}\Big(G(x+z,u,t)-G(x,u,t)\Big)\eta(z)dzdt+\int_0^{t_f}\frac{\partial G}{\partial u}du(t)\Big]>0 \\
&= \int_0^{t_f}\Big[\frac{\partial G}{\partial x}A(t)dt+\frac{1}{2}\frac{\partial^2 G}{\partial x^2}\sigma^2(t)+\frac{\partial G}{\partial u}du(t) \\
&+\lambda_J(x(t))\int_{\mathcal{Q}}\Big(G(x+z,u,t)-G(x,u,t)\Big)\eta(z)dzdt\Big]>0
\end{aligned}
\tag{2.4.11}
$$

G is a concave function of x, $\frac{\partial^2 G}{\partial x^2}\leq 0$. So,

$$
\begin{aligned}
V(u)-V(u^*) &= \int_0^{t_f}\frac{\partial G}{\partial x}A(t)dt+\int_0^{t_f}\frac{\partial G}{\partial u}du(t) \\
&+\lambda_J(x(t))\int_0^{t_f}\int_{\mathcal{Q}}\Big(G(x+z,u,t)-G(x,u,t)\Big)\eta(z)dzdt>0
\end{aligned}
\tag{2.4.12}
$$

$\square$

Theorem 2.4.4 *Let $u^*(t)$ be an admissible solution (control); then $u^*(t)$ is not a Nash equilibrium optimal solution only if there exists a piecewise continuous function $du(t)\in R^N$ such that the following inequality holds:*

$$
E\Big[\int_0^{t_f}\frac{\partial G}{\partial x}\Big(A(t)dt+dP(t)\Big)+\int_0^{t_f}\frac{\partial G}{\partial u}du(t)\Big]>0 \tag{2.4.13}
$$

Proof.

Necessity: Suppose $u^*(t)$ is unacceptable. Then, $V(u)>V(u^*)$, i.e.,

$$
E\int_0^{t}[G(x,u,t)-G(x^*,u^*,t)]dt>0
$$

Because $G(x,u,t)$ is a concave function of x,u, the following inequality holds.

$$
\frac{\partial G}{\partial x}\Big(x(t)-x^*(t)\Big)+\frac{\partial G}{\partial u}\Big(u(t)-u(t)^*\Big)\geq G(x,u,t)-G(x^*,u^*,t)
$$

Namely,

$$\int_0^{t_f} \frac{\partial G}{\partial x}\Big(x(t) - x^*(t)\Big) + \frac{\partial G}{\partial u}\Big(u(t) - u^*(t)\Big) \geq \int_0^{t_f} G(x,u,t) - G(x^*,u^*,t)dt$$

So,

$$\begin{aligned} & E\int_0^{t_f}\left[\frac{\partial G}{\partial x}\Big(A(t)dt + \sigma(t)dW(t)\Big) + dP(t) + \frac{\partial G}{\partial u}du\right] > 0 \\ & = E\int_0^{t_f}\left[\frac{\partial G}{\partial x}\Big(A(t)dt + dP(t)\Big) + \frac{\partial G}{\partial u}du\right] > 0 \end{aligned} \tag{2.4.14}$$

□

2.5 A Stochastic Differential-Jump Game Model for Brand Competition

A SDJG is used to model the brand competition problem. The Sethi and Lanchester framework is extended by additional jump events. The amplitudes of jumps are considered a uniform distributed and a normally distributed random variable, respectively.

In this section, we consider the "two companies" case. On the theoretical side, there is no question for this model to handle an N-player situation. However, multi-player games drive massive computations that make it very hard to solve these N-player problems when N is a very large case, for example, $N > 100$.

For two companies, we assume the summation of market share is equal to one, i.e., $x(t) + y(t) = 1$ by the normalization (Erickson (1995a, 1995b) and Prasad and Sethi (2004)). This equation assumes the total market share is constant. This assumption is a common condition and accepted by default. It is also the feature of a game model. This equation means the objective of competitors are conflict. When the market share of one company is increased, the market share of the other is decreased. A brand's market share depends not only on advertising, but also on price and the characteristics of the brand. The quality of products is fixed after they are manufactured. We consider only the advertising effort $u(t)$ as our control input variable, where $u(t) = (u_1(t), u_2(t))$.

$$
\begin{aligned}
V_1^*(x) &= \max_{u_1}\left\{E\int_0^\infty e^{-rt}[m_1x(t) - c_1u_1^2(t)]dt\right\}\\
V_2^*(y) &= \max_{u_2}\left\{E\int_0^\infty e^{-rt}[m_2y(t) - c_2u_2^2(t)]dt\right\}\\
dx(t) &= [\rho_1\sqrt{1-x(t)} - \rho_2\sqrt{x(t)}]dt + \sigma_1(t)dW(t) + dP(t)\\
x(t) + y(t) &= 1\\
x(t) &\geq 0\\
y(t) &\geq 0\\
\rho_1 &= \frac{\exp(\beta_{10} + u_1(t)\beta_1)}{\exp(\beta_{10} + u_1(t)\beta_1) + \exp(\beta_{20} + u_2(t)\beta_2)}\\
\rho_2 &= \frac{\exp(\beta_{20} + u_2(t)\beta_2)}{\exp(\beta_{10} + u_1(t)\beta_1) + \exp(\beta_{20} + u_2(t)\beta_2)}
\end{aligned}
\tag{2.5.1}
$$

where $x(t)$ and $y(t)$ are the sales rates (expressed as a fraction of the total market) at time t; $u_1(t), u_2(t)$ are advertising expenditure rates (efforts); r is the discount coefficient; and m_1, m_2, c_2, c_2 are profit coefficients and cost coefficients, respectively. ρ_1, ρ_2 are the response coefficients. δ is the coefficients of market share decay.

Based on Gihman and Skorohod (1972), as long as $u_i(x)$ and $\sigma(x)$ satisfy the Lipschitz conditions, and

$$\sigma(x) > 0, x \in (0,1), \sigma(0) = \sigma(1) = 0, \quad \text{and} \quad dP(t) = 0 \quad \text{at} \quad x = 1$$

x is almost surely $\in (0,1)$ because the drift is strictly positive at $x = 0$, and drift is strictly negative at $x = 1$.

$$\rho_1\sqrt{1-0} > 0, \quad \text{and} \quad -\rho_2 < 0$$

The HJB equation for each company is

$$
\begin{aligned}
rV_1^*(x) &= \max_{u_1}\Bigg\{ m_1 x - c_1 u_1^2 + \frac{\partial V_1^*}{\partial x}(\rho_1\sqrt{1-x} - \rho_2\sqrt{x}) \\
&\quad + 1/2\sigma_1^2 \frac{\partial^2 V_1}{\partial x^2} + \sum_z \lambda_J(x,u)(V_1^*(x+z) - V_1^*(x))\eta(z) \Bigg\} \\
rV_2^*(y) &= \max_{u_2}\Bigg\{ m_2(1-x) - c_2 u_2^2 + \frac{\partial V_2^*}{\partial x}(\rho_1\sqrt{1-x} - \rho_2\sqrt{x}) \\
&\quad + 1/2\sigma_1^2 \frac{\partial^2 V_2}{\partial x^2} + \sum_z \lambda_J(x,u)(V_2^*(x+z) - V_2^*(x))\eta(z) \Bigg\}
\end{aligned}
\tag{2.5.2}
$$

Here, x is $x(t)$ for simplification. We set

$$
g_1(x,u) = m_1 x - c_1 u_1^2, \quad g_2(x,u) = m_2(1-x) - c_2 u_2^2
$$

$$
f(x,u) = \rho_1\sqrt{1-x} - \rho_2\sqrt{x}
$$

After applying finite difference method, we get one Markov chain.

$$
\begin{aligned}
V_1^*(x) &= \max_{u_1}\Bigg\{ \sum p(x, x \pm h|u_1) V_1^*(x \pm h) + g_1(x,u)\triangle t + \mathcal{J}_1 \triangle t \Bigg\} \\
p(x, x \pm h|u_1) &= \frac{\sigma_1^2(t)/2 + hf^{\pm}(x,u)}{\sigma_1^2(t) + h|f(x,u)| + rh^2} \\
\triangle t &= \frac{h^2}{\sigma_1^2(t) + h|f(x,u)| + rh^2} \\
\mathcal{J}_1 &= \int_{\mathcal{Q}} \lambda_J(x,u) \Big[V_1^*(x+z) - V_1^*(x) \Big] \eta(z) dz
\end{aligned}
\tag{2.5.3}
$$

$$
\begin{aligned}
V_2^*(x) &= \max_{u_2}\Bigg\{ \sum p(x, x \pm h|u_2) V_2^*(x \pm h) + g_2(x,u)\triangle t + \mathcal{J}_2 \triangle t \Bigg\} \\
p(x, x \pm h|u_2) &= \frac{\sigma_1^2(t)/2 + hf^{\pm}(x,u)}{\sigma_1^2(t) + h|f(x,u)| + rh^2} \\
\triangle t &= \frac{h^2}{\sigma_1^2(t) + h|f(x,u)| + rh^2} \\
\mathcal{J}_2 &= \int_{\mathcal{Q}} \lambda_J(x,u) \Big[V_2^*(x+z) - V_2^*(x) \Big] \eta(z) dz
\end{aligned}
$$

where $n_p = 1, n_w = 1$, n_p and n_w are the number of Poisson processes and Brownian motion processes, respectively. λ_J is the rate of a jump process. z is the jump value, and h is the direction or step coefficient. $\mathcal{Q}$ is the jump amplitude domain.

$\eta(z)$ is the probability density of z.

Let us analyze the structure of this model. By applying the first order condition in the HJB equation (2.5.2), we have two equations:

$$
\begin{aligned}
0 &= 2c_1u_1(t) + \frac{\partial V_1^*}{\partial x}\left(\frac{\partial \rho_1}{\partial u_1}\sqrt{1-x} - \frac{\partial \rho_2}{\partial u_1}\sqrt{x}\right) \\
0 &= 2c_2u_2(t) - \frac{\partial V_2^*}{\partial x(t)}\left(\frac{\partial \rho_2}{\partial u_2}\sqrt{1-x} - \frac{\partial \rho_1}{\partial u_2}\sqrt{x}\right) \\
\frac{\partial \rho_1}{\partial u_1} &= \frac{\beta_1 \exp(\beta_{10} + u_1\beta_1 + \beta_{20} + u_2\beta_2)}{\Big(\exp(\beta_{10} + u_1\beta_1) + \exp(\beta_{20} + u_2\beta_2)\Big)^2} \\
\frac{\partial \rho_1}{\partial u_2} &= \frac{\beta_2 \exp(\beta_{20} + u_2\beta_2)}{\Big(\exp(\beta_{10} + u_1\beta_1) + \exp(\beta_{20} + u_2\beta_2)\Big)^2} \\
\frac{\partial \rho_2}{\partial u_1} &= \frac{\beta_1 \exp(\beta_{10} + u_1\beta_1}{\Big(\exp(\beta_{10} + u_1\beta_1) + \exp(\beta_{20} + u_2\beta_2)\Big)^2} \\
\frac{\partial \rho_2}{\partial u_2} &= \frac{\beta_2 \exp(\beta_{10} + u_1\beta_1 + \beta_{20} + u_2\beta_2)}{\Big(\exp(\beta_{10} + u_1\beta_1) + \exp(\beta_{20} + u_2\beta_2)\Big)^2}
\end{aligned}
\tag{2.5.4}
$$

Following Sethi (1983), we assume

$$
V_1(x) = a_1 + b_1x, \quad V_2(x) = a_2 + b_2(1-x) \tag{2.5.5}
$$

a_1, a_2, b_1, b_2 are constants. Combining equation (2.5.4) with (2.5.5), we get

$$
c_1u_1^*\beta_2b_2 + c_2u_2^*\beta_1b_1 = 0
$$

This is the meaning of a non-cooperative game. Basically, the effort of each player is in the opposite direction. The effort of player A will only benefit himself.

2.6 Numerical Examples

In this section, we present a number of examples to study the optimal control of market share with jump events and without jump events. Example one is a stochastic differential game. The other examples are stochastic differential-jump games.

Example one (Figure 2) is about a stochastic differential game model for brand competition. The system model is

$$
\begin{aligned}
V_1^*(x) &= \max_{u_1}\left\{E\int_0^\infty e^{-0.1*t}[100x(t) - 10u_1^2(t)]dt\right\} \\
V_2^*(y) &= \max_{u_2}\left\{E\int_0^\infty e^{-0.1*t}[200y(t) - 20u_2^2(t)]dt\right\} \\
dx(t) &= [\rho_1\sqrt{1-x(t)} - \rho_2\sqrt{x(t)}]dt + 0.2dW(t) \\
1 &= x(t) + y(t) \\
\rho_1 &= \frac{\exp(20u_1(t)+10)}{\sum\exp(\beta_{i0}+u_i(t)\beta_i)} \\
\rho_2 &= \frac{\exp(30u_2(t)+20)}{\sum\exp(\beta_{i0}+u_i(t)\beta_i)} \\
x(t) &\geq 0 \\
y(t) &\geq 0
\end{aligned}
$$

where $\sum\exp(\beta_{i0}+u_i(t)\beta_i) = \exp(20u_1(t)+10) + \exp(30u_2(t)+20)$. For instance, Figure (2) shows the profits and advertising efforts of companies A and B. In the following examples, the structure of figures follows the same sequence: profit, policy (advertising effort). In Figure (2), we know that the max profit of company A is 43, which is achieved when its market share is 0.92. When the market share of company A is 0.2, company A needs to increase its advertising effort from 0.01 to 0.4. When the market share of company A is 0.92, company A should decrease its advertising effort from 0.4 to 0.01.

The max profit of company B is 89.5, which is achieved at 0.82 of the whole market share. The advertising effort of company B is 0.02 except at the interval $[0.8, 1.0]$. The max advertising effort is around 0.14.

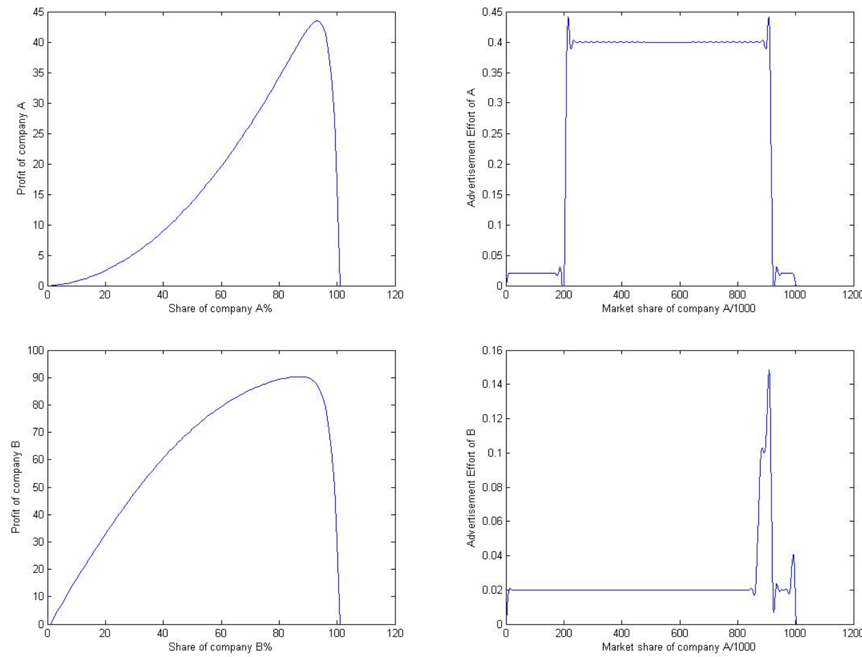

Figure 2: Stochastic differential game example: profits and efforts of companies.

Example two (Figure 3) is a stochastic deferential jump game. The system model is

$$V_1^*(x) = \max_{u_1}\left\{E\int_0^\infty e^{-0.1*t}[100x(t) - 10u_1^2(t)]dt\right\}$$
$$V_2^*(y) = \max_{u_2}\left\{E\int_0^\infty e^{-0.1*t}[200y(t) - 20u_2^2(t)]dt\right\}$$
$$dx(t) = [\rho_1\sqrt{1-x(t)} - \rho_2\sqrt{x(t)}]dt + 0.2dW(t) + dP(t)$$
$$1 = x(t) + y(t)$$
$$\rho_1 = \frac{\exp(20u_1(t)+10)}{\sum\exp(\beta_{i0}+u_i(t)\beta_i)}$$
$$\rho_2 = \frac{\exp(30u_2(t)+20)}{\sum\exp(\beta_{i0}+u_i(t)\beta_i)}$$

$\lambda_J = (1 + sin(t*0.5))/10$, the amplitude is a truncated normal random variable Norm(0.04, 0.01) in [-0.05,0.05].

In Figure (3), we know that the max profit of company A is 43, which is achieved when its market share reaches 0.92. When the market share of company A is 0.2, company A needs to increase its advertising effort from 0.01 to 0.4. When the market share of company A is 0.92, company A should decrease its advertising effort from 0.4 to 0.01. There are oscillations between $[0.8, 1.0]$.

The max profit of company B is 149, which is achieved when its the market share is 0.82. The advertising effort of company B is 0.02 except at the interval $[0.8, 1.0]$. The max advertisement effort is around 0.14.

We see that company A does not get the profit from the jump events. It seems that the jumps are too small for company A. But company B is sensitive to these jump events and gets the benefit of attracting customers from the increased market share of company A.

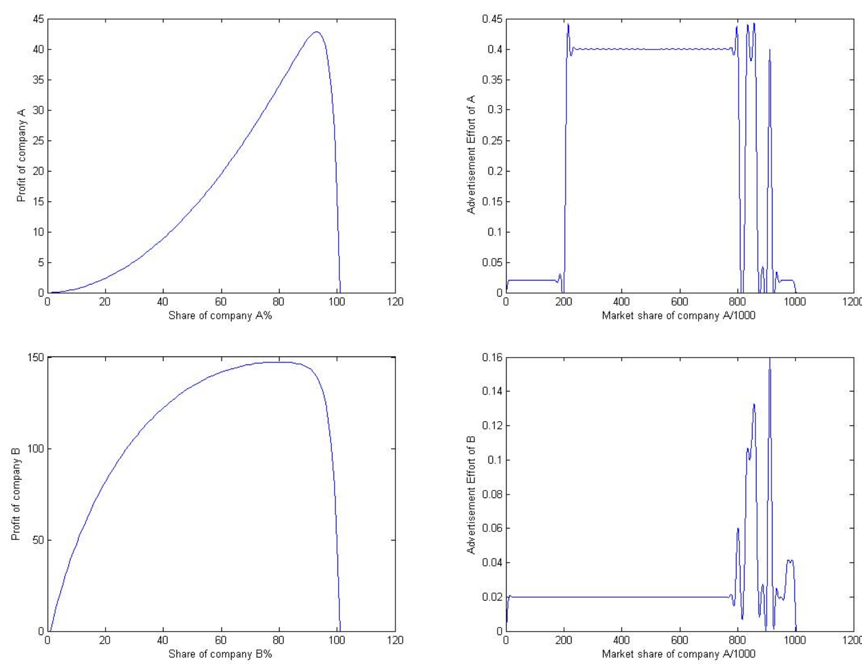

Figure 3: Profits and policy of companies A and B with normally distributed jump amplitude.

Example three (Figure 4) is a SDJG with uniformly distributed jump amplitude. The system model is

$$
\begin{aligned}
V_1^*(x) &= \max_{u_1}\left\{E\int_0^{\infty} e^{-0.1*t}[100x(t) - 10u_1^2(t)]dt\right\}\\
V_2^*(y) &= \max_{u_2}\left\{E\int_0^{\infty} e^{-0.1*t}[200y(t) - 20u_2^2(t)]dt\right\}\\
dx(t) &= [\rho_1\sqrt{1-x(t)} - \rho_2\sqrt{x(t)}]dt + 0.2dW(t) + dP(t)\\
1 &= x(t) + y(t)\\
\rho_1 &= \frac{\exp(20u_1(t)+10)}{\sum\exp(\beta_{i0}+u_i(t)\beta_i)}\\
\rho_2 &= \frac{\exp(30u_2(t)+20)}{\sum\exp(\beta_{i0}+u_i(t)\beta_i)}
\end{aligned}
$$

the $\lambda_J = (1 + sin(t * 0.5))/10$, the amplitude is a uniform random variable unif[-0.05, 0.05].

In Figure (4), we know that the max profit of company A is 60, which is achieved at market share 0.92. The advertising effort of company A is 0.4 except at the interval $[0.8, 1.0]$. The max advertising effort is around 0.45.

The max profit of company B is 149, which is achieved at market share 0.82. The advertising effort of company B is 0.02 except at the interval $[0.8, 1.0]$. The max advertising effort is around 0.14.

We see companies A and B do get the profit from the jump events at this time.

2.7 Discussion

We explore a stochastic dynamic duopoly game with jump events. We consider the feedback Markov solution. The solution satisfying HJB is the Nash equilibrium solution because we assume other players also use the Nash equilibrium soluation.

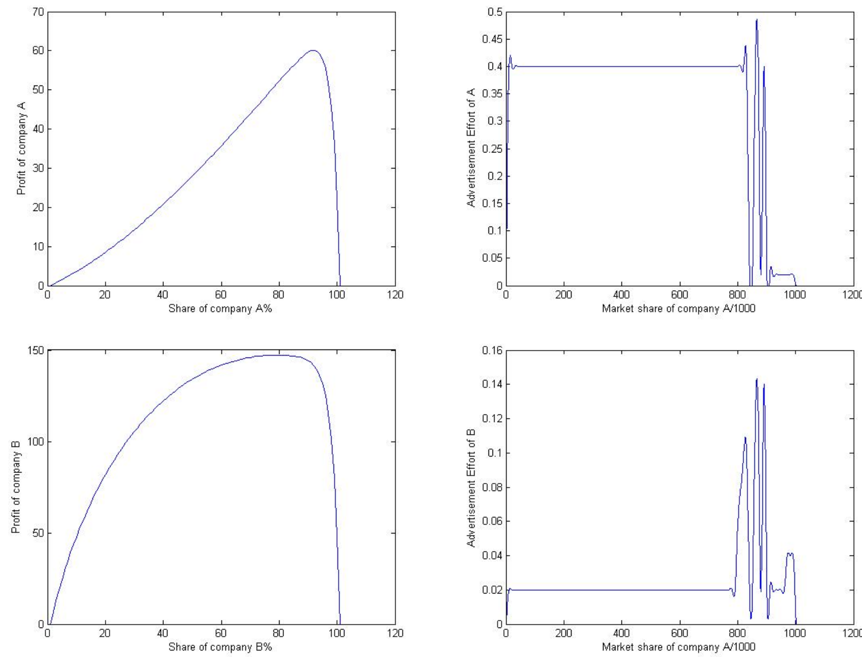

Figure 4: Profits and policy of companies A and B with uniformly distributed jump amplitude.

Our stochastic differential-jump game model extends the ability of the stochastic differential game model when there are jump disturbances in the market. The game model is developed from a static game model (mathematical programming) to a dynamic game model, namely, the differential game model that introduces the state evolution along time t. From a differential game model to a stochastic differential game model, the stochastic component, a Brownian motion is added to the state dynamic equation. In order to model the jumps/discrete events, the mark-time Poisson process is added in SDJG.

The state dynamic equation is extended by introducing jump events. The state dynamics are located between [0,1] by the absorbing boundary and reflective boundary condition. The HJB equation is also changed by adding an integration of the value function.

The Markov chain approximation method is preferred because of its good convergence properties. The method is developed from the finite difference method with a different way, namely, Markov decision process.

The numerical study shows some results of jump effects on optimal policies. The difficulty of computation originates from the curse of dimensionality of dynamic programming. Probably, the approximate dynamic programming method (Powell 2007) can reduce the computational burden.

In this chapter, we have defined and developed a new concept, a stochastic differential-jump game. This model enhances our ability to handle discrete state changes. Because not all system states can be modeled as continuous variables, this model may be useful, especially in marketing competition, when there is a powerful entrant to a balanced market. Our model follows the Sethi and Lanchester framework. The basic difference is that mark-time Poisson processes are added into state equations. Several examples in different situations were tested. To the best of our knowledge, these examples are the first applications of a stochastic differential-jump game model in market competition studies.

CHAPTER III

ROBUST HIERARCHICAL BAYESIAN LOGIT/PROBIT MODEL IN ANALYSIS OF MARKET HETEROGENEITY

Each customer is different, having a different age, income, education level, and culture characteristics that create heterogeneity. A hierarchical model is a suitable method for modeling heterogeneity. This type of model is used to provide a method to scrutinize detailed elements of a problem and provide a heterogeneity model (coefficients) for different customers.

A hierarchical model can combine the features of products with the characteristics of customers. The Markov Chain Monte Carlo (MCMC) simulation methodology makes the Bayesian technique more practical. MCMC has some advantages compared to maximum likelihood estimation (MLE) because MCMC realizes and draws inferences from those parameters samples.

Logit and probit models are simple and well-known statistical models. A hierarchical probit or logit model is a multi-level model in which the top level is the probit or logit, and the bottom level is a multilevel regression model. A robust hierarchical Bayesian logit/probit model is discussed in this chapter. Robustness also means that the model works very well with extreme outlier data because the long tail t distribution is used. Robustness makes estimation of those parameters more stable than that of a general maximum likelihood estimation (MLE) method.

This chapter includes (1) a robust hierarchical Bayesian logit/probit model and

(2) an experiment to validate the proposed robust model using a bank credit card choice data.

3.1 Introduction

Market heterogeneity is a phenomenon that is related to customer segmentation and target market selection. Simply, heterogeneity describes the difference among customers (Allenby et al. (1999)). Each customer has his/her own parameters even though we use the same model structure. Heterogeneity comes from demography, national culture, preferences, loyalties, etc. It is also related to the methods and processes by which a customer makes choices.

Logit and probit are well-known statistical models, but single-level logit and probit models have limited ability to deal with heterogeneity. A hierarchical model is a method that matches the classification or categorization process used by a human being. The advantage of this type of model is that it provides a method for scrutinizing detailed elements of a problem and provides different coefficients for different customers. A hierarchical model combines features of products with the characteristics of customers.

The MCMC methodology is a powerful computation technique for Bayesian analysis. There is an integral in the computation of a Bayesian posterior distribution that is difficult to compute. MCMC solves this problem and makes the Bayesian technique more practical.

There are two types of robust regression: the least squares error alternative and the parameter alternative. Huber introduced maximum likelihood type "M-estimation" for regression in 1973. This is a least squares alternative. Lange, Little, and Taylor (1989) introduced the parameters method, which is used in this chapter. The error is assumed to be a t distribution because t has a long tail to cover the outliers. Conversely, a normal distribution usually has a shrinkage effect. Because of the long tail, a t distribution can include some outlier data. The robust linear regression estimator assigns a different weight for each observation (sample) according to its estimated error.

The remainder of this chapter is organized as follows: Section (3.2) introduces several types of heterogeneity; argues why a hierarchical model is better for modeling heterogeneity; Section (3.3) discusses hierarchical model; Section (3.4) introduces the Markov Chain Monte Carlo simulation and its advantages; Section (3.5) reviews hierarchical Bayesian models for heterogeneity; Section (3.6) designs the robust hierarchical logit/probit model; Section (3.8) describes these experiments, and validation of our models.

3.2 Heterogeneity

Heterogeneity describes the differences among customers in a market. In the Merriam-Webster dictionary, the meaning of heterogeneity is the status of consisting of dissimilar or diverse ingredients or constituents. Researchers have studied several types of first level models with a regression at a lower level, for example, a first level normal linear regression model (Blattberg and George (1991)), a first level logit model (Allenby and Ginter (1995)), a first level probit model (McCulloch and Rossi (1994)), a first level Poisson model (Neelamegham and Chintagunta (1999)), and a first level generalized gamma distribution model (Allenby, Leone, and Jen (1999)). These models point out that there is a substantial degree of heterogeneity across units in various marketing data sets. Customer segmentation is also related to heterogeneity. For example, the price coefficient could be different for each group of customers because wealthier people are usually less price conscious. Common heterogeneity factors are reviewed below.

1) **Brand loyality/pereference heterogeneity**. Some customers are loyal to a specific brand. Preference segmentation is a widely researched issue in recent literature. For example, some customers have strong firm or country-origin loyalty in the auto market. There are two main methods to deal with this kind of heterogeneity. One method is to treat preference segmentation as an exogenous variable; the other is to treat preference segmentation as an unobserved variable distributed among different customers (which is thought to ignore heterogeneity in this method). Krishnamurthi and Raj (1988) treated preference heterogeneity as an average, and Guadagni and Little (1983) used exponential smoothing of past choices. Kamakura et al. (1991) used a continuous distribution with a random effects method. Jones and Landewehr (1988), Rossi and Allenby (1993), and Heckman and Singer (1984) used a discrete distributed or household-specific fixed effect.

2) **Price heterogeneity**. Some groups of customers are sensitive to price change. For instance, households with many children usually prefers a low price because they need a larger quantity of a particular product. The poor have greater budget (liquidity) constraints and may consequently buy products in small quantity, which results in higher per-unit retail prices. The groups with low incomes face higher retail prices than rich groups, particularly in developing countries (Rao (2000)).

3) **Structural heterogeneity**. Structural heterogeneity addresses the difference of a customer's choice processes. Different customers use different decision or choice

processes. This situation is under-researched compared to the preference segment. A common method for studying structural heterogeneity is a nested logit model. Currim et al. (1988) used nonparametric classifications. Kannan and Wright (1991) combined several nested logit models.

4) **Demographic heterogeneity**. Customers with the same demographic variables such as income, number of children, and age are grouped into the same segment. This research has been applied to retailer site choice (Ghosh and McLafferty (1987)), transportation method (Hauser, Koppelman, and Tybout (1981)), and market segmentation (Allenby and Rossi (1991)).

5) **Scale heterogeneity**. In survey research, survey practitioners have long noticed that respondents vary in their usage of scales: different patterns include using the middle of the scale or using the upper or lower bounds. Because there are large cultural or cross-country differences in scale usage, it is difficult to combine data across cultural or international boundaries. This different usage of scales, termed as "scale usage heterogeneity," imparts bias in many of the standard analyses conducted with survey data. The standard procedure for dealing with scale usage heterogeneity is to normalize data.

6) **Spatial heterogeneity**. The identification of geographic target markets is critical to the success of companies that are expanding internationally. The customer choice process depends on differences in cultures and nations. Country borders have traditionally been used to delineate such target markets. This is a "countries-as-segments" approach, even though there is an accelerating trend toward global market convergence. Researchers such as Sethi (1971), Helsen et al. (1993), and Kale (1995)

have used “country segmentation.” Some researchers have proposed using “cross-national segmentation” where country borders are ignored and consumers in different countries are grouped into cross-national segments, based on their similarities in needs (Kamakura et al. (1993), Yavas et al. (1992)).

Below is an example of the heterogeneity model. The following is a logit model:

$$p_{\hat{i}i} = \frac{\exp(x_i'\beta_{\hat{i}})}{\sum_j \exp(x_j'\beta_{\hat{i}})}$$

where $x_i = (x_{i1}, ..., x_{ik})'$ are the choice attributes, and $\beta_i = (\beta_{i1}, ..., \beta_{ik})'$ are the parameters associated with these choice attributes. If coefficients $\beta_{\hat{i}}$ are the same for each customer, the logit model is a homogenous model. If coefficients β_i are different for each customer, the logit model is a heterogenous model (Jones et al. (1988)).

The formal argument is why heterogeneous models work better than homogeneous models. Assume that we use a heterogeneous design so that respondent $\hat{i}$ is given the parameters $\beta_{\hat{i}}$, for $\hat{i} = 1, ..., N$, and the full parameters consist of the collection of subparameters $\beta = \beta_{\hat{i}}; \hat{i} = 1, ..., N$. Here, each respondent would have his or her own parameters. We obtain the optimal heterogeneous model by maximizing $Likelihood(\beta)$ over the N subparameters. The optimization problem can be formally represented as follows:

$$\max Likelihood(\beta_1, \beta_2, ..., \beta_N)$$

Note that we obtain the homogeneous model by maximizing $Likelihood(\beta)$ with respect to β, which is the same as maximizing $Likelihood(\beta_1, ..., \beta_N)$ with respect to $\beta_1, ..., \beta_N$ under the restriction that $\beta_1 = \beta_2, \cdots, = \beta_N$. So, the homogenous model is a constrained optimization problem. A heterogenous model is an unconstrained optimal problem compared with a homogenous model. Thus the result of a heterogenous model is better than that of a homogenous model.

3.3 Hierarchical Model

A hierarchical model is a multi-level model that fits the nature of classified and categorized data. It is very suitable for modeling market heterogeneity. Many marketing data sets have a hierarchical or cluster structure. For example, we first observe the choice data, "0" or "1" for a brand. Then we look for why the customer chose this brand. Is the choice because of a specific customer or because of a specific brand?

Before using the hierarchical model, an alternative approach for heterogeneity is to specify a random effects model. Examples include Heckman and Singer (1984), Kamakura and Russell (1989), Chintagunta et al. (1991), and Gonul and Srinivasan (1993). In the random effects model, individual household level parameters are viewed as draws from a super-population distribution, which is called the random effects or a mixed distribution.

Hierarchy describes a model consisting of units grouped in different levels. In these models, there are a few lower level units within each higher level unit. Especially in some complex cases, this kind of logit or probit model with a binary (0,1) response is no longer adequate. Rodriguez and Goldman (1995) illustrated this hierarchical model. Goldstein and Rasbash (1996) also described this multi-level model.

For example, the hierarchy is

Likelihood: $p(y_i \mid \theta_i)$
First-stage prior: $p(\theta_i \mid \tau)$
Second-stage prior: $p(\tau \mid x_i)$

The y_i, x_i are the output and input information, respectively. θ_i, τ are the parameters of the hierarchical model.

The joint posterior for the hierarchical model is given by

$$p(\theta_1, ..., \theta_m, \tau | y_1, ..., y_m, x_1, ..., x_m) \propto \Pi_{i=1}^m p(y_i|\theta_i)p(\theta_i|\tau)p(\tau|x_i).$$

In a hierarchical model, the prior induced on the unit-level parameters is not an independent prior. The unit-level parameters are conditionally independent :

$$p(\theta_1, \theta_2, ...\theta_m | x_1, ..., x_m) = \int \Pi_i p(\theta_i|\tau)p(\tau|x_i)d\tau$$

In the sequel, we present an overview of a three-level logit model: A three-level logit model is a two-level logit model with a lower level regression for each β_i, the parameter associates with choice attributes.

First level: Here, observation $y_{\hat{i}i}$ indicates if brand i is chosen. $y_{\hat{i}i}$ is a Bernoulli random variable with success probability $y_{\hat{i}i}|p_{\hat{i}i}$, $y_{\hat{i}i} = 1$ with probability $p_{\hat{i}i}$. $y_{\hat{i}i} = 1$ means brand i is chosen by the respondent $\hat{i}$. $y_{\hat{i}i} = 0$ means the brand i is not chosen by the respondent $\hat{i}$.

Second level: It is a logit model. $x_i = (x_{i1}, ..., x_{ik})'$ are the choice attributes, and $\beta_{\hat{i}} = (\beta_{\hat{i}1}, ..., \beta_{\hat{i}k})'$ are the parameters associated with these choice attributes.

$$p_{\hat{i}i} = \frac{\exp(x_i'\beta_{\hat{i}})}{\sum_j \exp(x_j'\beta_j)}$$

Third level: A linear regression is used.

$$\beta_{\hat{i}} = \Gamma z_{\hat{i}}' + \varepsilon$$

where $z_{\hat{i}} = (z_{\hat{i}1}, ..., z_{\hat{i}m})$, $\varepsilon \sim Norm(0, \sigma^2)$. The three-level probit model is a two-level probit model with a regression or distribution for $\beta_{\hat{i}}$. The third level is a linear regression for $\beta_{\hat{i}}$, where $\beta_{\hat{i}} = \Gamma z'_{\hat{i}} + \varepsilon$.

3.4 Markov Chain Monte Carlo

The general computational difficulty facing Bayesians is that various integrals of functions with respect to the posterior distribution must be computed. Since these integrals can be written as the posterior expectation of a function of the parameters, simulation methods seem natural for approximation. For example, if we could make independent identically distributed (i.i.d) draws from the posterior, we could simply approximate the integrals by the sample mean

$$\mathbb{I} = E_{\theta|y}[f(\theta)] = \int f(\theta)p(\theta|y)d\theta \quad \text{and} \quad \hat{\mathbb{I}} = \frac{1}{n}\sum_{i=1}^{\bar{N}} f(\theta_i)$$

where $\bar{N}$ is the number of sampling, and θ is the parameter that f bases on. Instead of using i.i.d. draws, another approach could be to construct a Markov chain with the posterior as its equilibrium distribution. In practice, this means specifying a transition density that produces a sequence of θ draws. θ_i is a draw from $p(\theta_i|\theta_{i-1})$ given θ_{i-1}.

MCMC has some advantages in computation: (1) MCMC has the ability to handle many different types of variables, and inferences are valid for all parameters; (2) missing responses are treated in a right way and missing data can be drawn and replaced from the posterior distribution; and (3) hierarchical models are naturally implemented by MCMC; many types of data used in marketing research are generated by a hierarchical process.

The MCMC method is very suitable for developing a hierarchical model. The hierarchical structure of panel data is ideally suited for the MCMC method. There are two well-known algorithms to implement a Markov Chain Monte Carlo (MCMC) simulation: the Metropolis-Hastings algorithm (Metropolis et al. (1953)) and the Gibbs sampler algorithm (Geman et al. (1984)).

Robust regression estimation has recently attracted much attention, but the general usage of regression is not a robust estimator because the error distribution is assumed to be a normal distribution. Actually, maximum likelihood estimators (MLE) are very sensitive to the assumptions of the statistical model. For instance, consider the assumption that the sample mean and variance are truly normally distributed. If the assumption is true for sample data, then the sample mean and variance are correctly estimated. However, these estimators will change significantly when there are outliers. Because these estimators are the average of the data and the squared deviations from the mean, respectively, a single outlier may drive the sample mean and variance to an erroneous value when it is far enough removed from other data. Actually, extreme outliers are nearly impossible to be fitted with an exactly normal distribution. We usually approximate the data by a t distribution with heavier tails instead of a normal distribution.

3.5 Review of Previous Models

In the marketing science literature, there are nested logit models, mixture normal models, hierarchical models, and structural models. Here, we review some hierarchical models that relate to our work.

Hierarchical Model. A hierarchical logit model is given by McCulloch and Rossi

(1994).

$$P(\tilde{Y}_{\hat{i}j} = 1) = \frac{\exp(y_j)}{\sum_k \exp(y_k)},$$
$$y_{\hat{i}j} = X'_{\hat{i}j}\beta_{\hat{i}} + \varepsilon_{\hat{i}j}, \quad \varepsilon_{\hat{i}} \sim Norm(0, \sigma^2),$$
$$\beta_{\hat{i}} = \Delta' z_{\hat{i}} + v_{\hat{i}}, \quad v_{\hat{i}} \sim Norm(0, V_\beta).$$

Here, $\tilde{Y}_{\hat{i}j} = 1$ when brand j is chosen. $y_{\hat{i}j}$ is the latent variable. $z_{\hat{i}}$ is a vector of variables that explain heterogeneity as cross-unit differences. Typically, we use those various demographic or market characteristics that connect differences in brand preference or marketing sensitivities.

Allenby and Ginter (1995) gave another hierarchical model :

$$P(\tilde{Y}_{\hat{i}j} = 1) = \frac{\exp(y_{\hat{i}j})}{\sum_k \exp(y_{\hat{i}k})},$$
$$y_{\hat{i}j} = x'_{\hat{i}j}(\alpha + \beta_{\hat{i}}) + \varepsilon_{\hat{i}j},$$
$$\beta_{\hat{i}} = \Gamma z_{\hat{i}} + \zeta_{\hat{i}} \quad .$$

where

$y_{\hat{i}j}$: the latent variable of profile j evaluated by the respondent $\hat{i}$,

$x_{\hat{i}j}$: vector of independent variables associated with profile j for respondent $\hat{i}$,

α : fixed-effects coefficient that are assumed constant across respondents,

$\beta_{\hat{i}}$: respondent-specific coefficients that modify α,

$\varepsilon_{\hat{i}j}$: i.i.d norm error term $Norm(0, \sigma^2)$,

$\zeta_{\hat{i}}$: i.i.d $Norm(0, D)$,

Γ : a matrix of coefficients that relates $\beta_{\hat{i}}$ the value of $z_{\hat{i}}$,

$z_{\hat{i}}$: the vector of variables that account for observed heterogeneity.

Yang and Allenby (2003) proposed an autoregression model to reflect the possible interdependent effects of each customer because of social relationship and information networks:

$$
\begin{aligned}
P(\tilde{Y}_i = 1) &= P(y_i > 0),\\
y_i &= x_i'\beta + \varepsilon_i + \theta_i,\\
\theta &= \rho W \theta + \mu,\\
\varepsilon &= Norm(0, \bar{I}),\\
\mu &\sim Norm(0, \sigma^2).
\end{aligned}
$$

where ρW is a matrix that specifies the interdependence network, which is a coefficient that measures the influence of the network. $\theta = (\theta_1, ..., \theta_k)$ are the autoregression parameters.

Mixture Normal Model. The elemental block of a mixture of normal extensions of the prior pdf is the normal model. A mixture of normal models (Rossi et al. 2005) can be written as

$$
p(\theta|\overline{\theta}_1, ...\overline{\theta}_k, V_1, ..., V_k) = r_1\phi(\theta|\overline{\theta}_1, V_1) + r_2\phi(\theta|\overline{\theta}_2, V_2)+, ..., +r_k\phi(\theta|\overline{\theta}_k, V_k),
$$
$$
\sum_i r_i = 1.
$$

The mixture of normal models provides a great deal of flexibility for pdf approximation, especially for the long tail distribution.

Nested Logit Model. Moore and Lehmann (1989) and Jeffrey (1986) used those nested multinomial logit models. McFadden (1986), Ben-Arkiva and Lerman (1985) took into account the structural market as well. The structure of nested logit model is

$$
p_{ij} = \frac{e^{X_j\beta_i/\alpha_k}(\sum_{j\in B_k} e^{X_j\beta_i/\alpha_k})^{\alpha_k-1}}{\sum_{l=1}^{K}(\sum_{j\in B_l} e^{X_j\beta_i/\alpha_l})^{\alpha_l}}
$$

the $0 \leq \alpha_k \leq 1$ is a measure of the degree of independence in unobserved utility among alternatives in nest k. B_k represents subset, namely, nest k.

3.6 Bayesian Robust Regression

Robust statistics can work well to give useful results even though a certain specified model assumption is incorrect or when the assumed model error distribution is inappropriate. This class of statistics is useful when outliers (observations far from the bulk of the data) are present, for example, if error distribution has heavier tails such as a t distribution, Cauchy, or Double exponential distribution.

3.6.1 Some Important Probability Distributions

Before we talk about the robust hierarchical probit and logit models, we review some important probability distributions used in the robust regression.

(1) t distribution: Let Z be a $Norm(0,1)$ and $\chi^2(\nu)$ be a chi-square variable with ν degrees of freedom. If Z and $\chi^2(\nu)$ are independent, then the random variable

$$\bar{T} = \frac{Z}{\sqrt{\chi^2(\nu)/\nu}}$$

has the probability density function (pdf)

$$f(x) = \frac{\hat{\Gamma}[(\nu+1)/2]}{\sqrt{\pi\nu}\hat{\Gamma}(\nu/2)} \frac{1}{[(\frac{x^2}{\nu}+1)]^{(\nu+1)/2}}, -\infty < x < \infty$$
$$\hat{\Gamma}(x) = \int_0^\infty s^{(x-1)} e^{-s} ds,$$
$$\hat{\Gamma}(n) = n! \quad .$$

(2) The pdf of $\chi^2(\nu)$ is

$$f(x) = \frac{1}{2^{\nu/2}\hat{\Gamma}(\frac{\nu}{2})} x^{(\nu/2)-1} e^{-x/2}, x > 0.$$

(3) The pdf of Gamma distribution $Gamma(x; r, \lambda)$ is

$$f(x) = \frac{\lambda}{\hat{\Gamma}(r)} (\lambda x)^{r-1} e^{-\lambda x}, x > 0.$$

The Gamma distribution $\mathcal{G}amma(\nu/2, \nu/2)$ and $\chi^2(\nu)/\nu$ have the same probability density function.

Proof.

The probability density function of $\mathcal{G}amma(\nu/2, \nu/2)$ is given by

$$f(x|\nu) = \frac{(\nu/2)^{\nu/2} x^{\nu/2-1} e^{-x\nu/2}}{\hat{\Gamma}(\nu/2)}$$

The probability density function of $\chi^2(\nu)/\nu$, assumes $y \sim \chi^2(\nu)$, $x = y/\nu$, so the the probability density function of x is

$$\begin{aligned} f(x) &= \frac{1}{2^{\nu/2}\hat{\Gamma}(\frac{\nu}{2})}(\nu x)^{(\nu/2)-1} e^{-\nu x/2} * \frac{dy}{dx} \\ &= \frac{(\nu/2)^{\nu/2} x^{\nu/2-1} e^{-x\nu/2}}{\hat{\Gamma}(\nu/2)}. \end{aligned}$$

□

3.6.2 Bayesian Point Estimation

Lange, Little, and Taylor (1989) provided a robust parameter estimation in a regression model using a t distribution. Conditional on x_i and w_i, the y_i are independently distributed and,

$$y_i|\underline{X} \sim Norm(x_i^{'}\beta, \sigma^2/w_i)$$

where $\beta = (\beta_1, ..., \beta_k)^{'}$ is a vector of unknown parameters, $w = diag(w_1, ..., w_n)^{'}$ is a vector of unknown parameters, n is the number of observation, and σ is also an unknown parameter. N is a normal random variable. It is sometimes more useful to write this model as

$$y_i = x_i^{'}\beta + \varepsilon_i, \quad \varepsilon_i \sim Norm(0, \sigma^2/w_i), \quad (i = 1, ..., n).$$

or in vector formation.

$$y = \underline{X}\beta + \varepsilon, \quad var(\varepsilon) = \sigma^2/w, \quad w = diag(w_1, ..., w_n).$$

The likelihood function of this model is :

$$L(y, x; \beta, \sigma, w) = \sigma^{-n}\Pi_{i=1}^{n} w_i^{1/2} exp\left[-\sum_{i=1}^{n} w_i(y_i - x_i^{'}\beta)^2/2\sigma^2\right].$$

If the prior probability density function is $\pi(\beta, \sigma) = \sigma^{-1}$ and w_i is $\mathcal{G}amma(\nu/2, \nu/2)$, then

$$\Pi_i^n \pi(w_i) = (\nu/2)^{n\nu/2}\left[\frac{\Pi_{i=1}^{n} w_i^{(\nu+2)/2} exp(-\nu w_i/2)}{\hat{\Gamma}(\frac{\nu}{2})^n}\right]$$

So the likelihood function for the Bayesian posterior density kernel is

$$L(\beta, \sigma, w; y, x) \propto (\nu/2)^{n\nu/2}\frac{\sigma^{-(n+1)}\Pi_{i=1}^{n} w_i^{(\nu+3)/2} exp\{-\sum_{i=1}^{n}[\sigma^{-2}(y_i - x_i^{'}\beta)^2 + v]w_i/2\}}{\hat{\Gamma}(\frac{\nu}{2})^n}.$$

3.6.3 Posterior Density Distribution of β and σ

(a) First, we consider the multiple regression case. As the posterior density of β, σ by setting the x_i, y_i, w_i, ν_0, the posterior density of β and σ is proportional to $exp\left[-\frac{1}{2\sigma^2}(\beta - \tilde{\beta})'(\underline{X}^{'}\underline{X} + A)(\beta - \tilde{\beta})\right] exp\left[-\frac{(\nu_0 s_0^2 + ns^2)}{2\sigma^2}\right]$.

$$exp\left[-\sum_{i=1}^{n}\sigma^{-2}(y_i - x_i^{'}\beta)^2 w_i/2\right] = exp\left[-(y_i\sqrt{w_i} - x_i^{'}\sqrt{w_i}\beta)/2\sigma^2\right]$$

We define $\underline{X} = (x_1\sqrt{w_1}, ..., x_i\sqrt{w_i}, ..., x_n\sqrt{w_n})$, and $y = (y_1\sqrt{w_1}, ..., y_i\sqrt{w_i}, ..., y_n\sqrt{w_n})$.

$$
\begin{aligned}
p(\beta,\sigma^2|y,x) &\propto p(y|X,\beta,\sigma^2)p(\beta|\sigma^2)p(\sigma^2) \\
&\propto exp\left[-(y-\underline{X}\beta)^{'}(y-\underline{X}\beta)/2\sigma^2\right] \\
&\times exp\left[-(\beta-\bar{\beta})^{'}A(\beta-\beta)/2\sigma^2\right] \\
&\times exp\left[-\nu_0 s_0^2/2\sigma^2\right] \\
&\propto exp\left[-\frac{1}{2\sigma^2}(\beta-\tilde{\beta})^{'}(\underline{X}^{'}\underline{X}+A)(\beta-\tilde{\beta})\right] exp\left[-\frac{(\nu_0 s_0^2+ns^2)}{2\sigma^2}\right].
\end{aligned}
$$

So

$$
\beta|(\sigma,w) \sim Norm[\tilde{\beta},\sigma^2(\underline{X}'\underline{X}+A)^{-1}], \quad \sigma^2|y \sim \frac{\nu_1 s_1^2}{\chi^2(\nu_1)},
$$

and the prior distribution of β is $Norm(\bar{\beta},A)$,

$$
\begin{aligned}
\nu_1 &= \nu_0+n, s_1^2 = \frac{\nu_0 s_0^2+ns^2}{\nu_0+n}, \\
ns^2 &= (y-\underline{X}\tilde{\beta})^{'}(y-\underline{X}\tilde{\beta})+(\tilde{\beta}-\bar{\beta})^{'}A(\tilde{\beta}-\bar{\beta}), \\
\tilde{\beta} &= (\underline{X}^{'}\underline{X}+A)^{-1}(\underline{X}y+A\bar{\beta}).
\end{aligned}
$$

n is the observation number. $w = diag\{w_1,...,w_n\}$, $\beta = \{\beta_1,...,\beta_n\}$, and $x = \{x_1,...,x_n\}$.

(b) In the multivariate regression case

$$
\bar{Y}_j = \bar{X}_j\mathcal{B} + Norm(0,\Sigma/w_j)
$$

the dimensions of $\bar{Y},\bar{X}$ are $(n\times m)$, $(n\times k)$, $\bar{Y} = (\bar{Y}_1,...,\bar{Y}_j,...,\bar{Y}_n)'$, $\bar{Y}_j = (\bar{y}_{j1},...,\bar{y}_{jm})$, $\bar{X} = (\bar{X}_1,...,\bar{X}_j,...,\bar{X}_n)'$, $\bar{X}_j = (\bar{x}_{j1},...,\bar{x}_{jk})$,

$$
\bar{Y}_j\sqrt{w_j} = \bar{X}_j\sqrt{w_j} + Norm(0,\Sigma)
$$

We set $\bar{Y}_j\sqrt{w_j} = Y_j$, $\bar{X}_j\sqrt{w_j} = X_j$. So we have $Y = (\bar{Y}_1\sqrt{w_1},...,\bar{Y}_j\sqrt{w_j},...,\bar{Y}_n\sqrt{w_j})'$, $Y_j = \sqrt{w_j}(\bar{y}_{j1},...,\bar{y}_{jm})$, $\underline{X} = (\bar{X}_1\sqrt{w_1},...,\bar{X}_j\sqrt{w_j},...,\bar{X}_n\sqrt{w_n})'$, $X_j = \sqrt{w_j}(\bar{x}_{j1},...,\bar{x}_{jk})$.

Assume the $\mathcal{B}$ and Σ prior distribution is

$$p(\Sigma, \mathcal{B}) \propto |\Sigma|^{-(\nu_0+m+1)/2} \exp\left[-\frac{1}{2} tr\Sigma^{-1}V_0\right].$$

where ν_0, m, and V_0 are the prior parameters. $|\Sigma|$ is the determinant of Σ. By the optimal least squares estimation, we have

$$\hat{\mathcal{B}} = (\underline{X}'\underline{X})^{-1}\underline{X}'Y.$$

The posterior distribution has three levels. The following are the exact results of Rossi et al. (2005).

$$\begin{aligned}
p(\Sigma, \mathcal{B}|Y, \underline{X}, w) &\propto |\Sigma|^{-(\nu_0+m+1)/2} \exp\left[tr(-\frac{1}{2}V_0\Sigma^{-1})\right] \\
&\times |\Sigma|^{-k/2} \exp\left[tr(-\frac{1}{2}(\mathcal{B}-\bar{\mathcal{B}})'A(\mathcal{B}-\bar{\mathcal{B}})\Sigma^{-1})\right] \\
&\times |\Sigma|^{-n/2} \exp\left[tr(-\frac{1}{2}(Y-\underline{X}\mathcal{B})'(Y-\underline{X}\mathcal{B})\Sigma^{-1})\right] \\
&\propto |\Sigma|^{-(\nu_0+m+1)/2} \exp\left[tr(-\frac{1}{2}V_0\Sigma^{-1})\right] \\
&\times |\Sigma|^{-(k+n)/2} \exp\left[tr((-\frac{1}{2}(\mathcal{B}-\bar{\mathcal{B}})'A(\mathcal{B}-\bar{\mathcal{B}}) - \frac{1}{2}(Y-\underline{X}\mathcal{B})'(Y-\underline{X}\mathcal{B}))\Sigma^{-1})\right]. \\
&\propto |\Sigma|^{-(\nu_0+m+n+1)/2} \exp\left[tr(-\frac{1}{2}(V_0+S)\Sigma^{-1}))\right] \\
&\times |\Sigma|^{-k/2} \exp\left[tr(-\frac{1}{2}(\mathcal{B}-\tilde{\mathcal{B}})'(\underline{X}'\underline{X}+A)(\mathcal{B}-\tilde{\mathcal{B}})\Sigma^{-1})\right]
\end{aligned}$$

where

$$\tilde{\mathcal{B}} = (\underline{X}'\underline{X}+A)^{-1}(\underline{X}'\underline{X}\hat{\mathcal{B}}+A\bar{\mathcal{B}}),$$

$$S = (Y-\underline{X}\tilde{\mathcal{B}})'(Y-\underline{X}\tilde{\mathcal{B}}) + (\tilde{\mathcal{B}}-\bar{\mathcal{B}})'A(\tilde{\mathcal{B}}-\bar{\mathcal{B}}).$$

where $\bar{\mathcal{B}}$ is the prior estimated parameter and A is the variance matrix of $\mathcal{B}$ conditional on Σ. $trA = \sum_i^n a_{ii}$ is the trace of a square $n \times n$ matrix A. m is the dimension of Σ, k is the dimension of $\mathcal{B}$, and n is the number of observations of Y.

The posterior distribution of Σ matches the result of Anderson (1994). If A has the Wishart distribution $W(\Sigma, n)$ and Σ has the a priori inverse Wishart distribution $IW(\Psi, m)$, then the conditional distribution of Σ is $IW(A + \Psi, n + m)$.

3.6.4 Posterior Density Distribution of w and ν

The posterior distribution of w is conditional on β and σ; the conditional posterior density of w_i is proportional to

$$w_i^{(\nu-3)/2} exp[-(\sigma^{-2}(y_i - \beta_i x_i)^2 + \nu)w_i/2].$$

So, by Geweke (1993),

$$(\sigma^{-2}(y_i - \beta_i x_i)^2 + \nu)w_i|(\beta, \sigma) \sim \chi^2(\nu + 1)$$

Since the $\mathcal{G}amma$ distribution is the conjugate prior distribution,

$$w_i|y, x, \beta, \nu \sim \mathcal{G}amma(\frac{\nu + n}{2}, \frac{\nu + \sum_{i=1}^{n}(y_i - \beta_i x_i)^2}{2}).$$

The distribution of freedom ν was discussed by Geweke (1993) and Liu (1995). The following are some theoretical results used in hierarchical models.

3.6.5 Two Level Regressions

The following theories are the result of Bayesian analysis for a hierarchical regression model.

Theorem 3.6.1 *(Lindley and Smith (1972)) Given* θ_1*,*

$$y \sim Norm(A_1\theta_1, C_1)$$

and given θ_2

$$\theta_1 \sim Norm(A_2\theta_2, C_2),$$

then the distribution of y *is*

$$y \sim Norm(A_1A_2\theta_2, C_1 + A_1C_2A_1^{'}).$$

the condition distribution of θ_1 *on* y *is* $Norm(Bb, B)$ *with*

$$B^{-1} = A_1^{'}C_1^{-1}A_1 + C_2^{-1}, \quad b = A_1^{'}C_1^{-1}y + C_2^{-1}A_2\theta_2.$$

Using the above theorem,

$$y_{\hat{i}j} = x_{\hat{i}j}^{\prime}\beta_{\hat{i}} + \varepsilon_{\hat{i}}, \quad y_{\hat{i}j} \sim Norm(x_{\hat{i}j}^{\prime}\beta_{\hat{i}}, \sigma^2),$$
$$\beta_{\hat{i}} = \Gamma z_{\hat{i}} + \xi_{\hat{i}}, \quad \beta_{\hat{i}} \sim Norm(\Gamma z_{\hat{i}}, \xi_{\hat{i}}).$$

Draw $\beta_{\hat{i}}$, $\hat{i} = 1, ..., M$(one respondent at a time)

$$\beta_{\hat{i}}|\{x, y, Z, \Gamma, \sigma^2, \Sigma, w_{\hat{i}}\} \sim Norm(\beta_{\hat{i}}^*, B_{\hat{i}})$$

Using the above theorem and setting $\theta_1 = \beta_{\hat{i}}$, $C_1 = \sigma^2$, $A_1 = x_{\hat{i}j}$, and $\theta_2 = \Gamma$, $C_2 = \Sigma/w_{\hat{i}}$, $A_2 = z_{\hat{i}}$. Where

$$B_{\hat{i}} = ((\Sigma/w_{\hat{i}})^{-1} + \sigma^{-2}(\sum_j x_{\hat{i}j}x_{\hat{i}j}^{\prime}))^{-1},$$
$$\beta_{\hat{i}}^* = B_{\hat{i}}((\Sigma/w_{\hat{i}})^{-1}\Gamma z_{\hat{i}} + \sigma^{-2}(\sum_j x_{\hat{i}j}x_{\hat{i}j}^{\prime})\hat{\beta}_{\hat{i}}),$$
$$\hat{\beta}_{\hat{i}} = (\sum_j x_{\hat{i}j}x_{\hat{i}j}^{\prime})^{-1}(\sum_j x_{\hat{i}j}y_{\hat{i}j}).$$

3.7 Robust Hierarchial Bayesian Logit/Probit Model

Lange, Little, and Taylor (1989) gave a robust parameter estimation in a regression using a t distribution. In the newly developed model, a t distribution is used in the bottom level of the hierarchical logit/probit model.

$$\beta_i \sim t(\Gamma z_i, \Sigma, \nu)$$

First level: observation $y_{\hat{i}i}$ is modeled as a Bernoulli random variable with success probability: $y_{\hat{i}i}|p_{\hat{i}i}$, $y_{\hat{i}i} = 1$ with probability $p_{\hat{i}i}$.

The second level is a logit model:

$$p_{\hat{i}i} = \frac{\exp(x_i'\beta_{\hat{i}})}{\sum_j \exp(x_j'\beta_{\hat{i}})},$$

or a probit model

$$p_{\hat{i}i} = \Phi(x_i'\beta_{\hat{i}}) \quad .$$

The third level is a linear regression:

$$\beta_{\hat{i}} = \Gamma z_{\hat{i}} + \varepsilon.$$

where ε is $t(u, \Sigma, \nu_1)$ (Liu 1996). To estimate the $t(u, \Sigma, \nu_1)$, a normal/independent random variable $\bar{T} = \mu + V/\sqrt{w}$ is used, where V is the normal random variable $N(0, \Sigma)$. w is a positive independent random variable, which has a distribution $f(w|\nu_1)$, and a gamma distribution $\mathcal{G}amma(\nu_1/2, \nu_1/2)$. So

$$\beta_{\hat{i}}\sqrt{w_{\hat{i}}} = \Gamma z_i \sqrt{w_{\hat{i}}} + V$$

After reviewing the robust Bayesian estimation theory in the above section, we merge this method with the hierarchical framework and call the resulting model a robust hierarchical model. We need to assume the prior distribution first:

$$
\begin{aligned}
&\text{Brand/profile } j \text{ is chosen with probability } p_j, \\
&p_j = logit(x'_{\hat{i}j}\beta_{\hat{i}}) \quad \text{or} \quad p_j = \Phi(x'_{\hat{i}j}\beta_{\hat{i}}), \\
&y_{\hat{i}j} = x'_{\hat{i}j}\beta_{\hat{i}} + \varepsilon_{\hat{i}j}, \\
&\beta_{\hat{i}} = \Gamma z_{\hat{i}} + \xi_{\hat{i}}, \\
&\varepsilon_{\hat{i}j} \sim N(0, \sigma^2), \\
&\xi_{\hat{i}} \sim Norm(0, \Sigma/w_{\hat{i}}), \\
&w_{\hat{i}} \sim \mathcal{G}amma(\nu_1/2, \nu_1/2).
\end{aligned}
\tag{3.7.1}
$$

The following are definitions:

$y_{\hat{i}j}$: the latent variable of profile j evaluated by the respondent $\hat{i}$;

$x_{\hat{i}j}$: vector of independent variables associated with profile j for respondent $\hat{i}$;

$\beta_{\hat{i}}$: respondent i-specific coefficients, $\beta_{\hat{i}} = (\beta_{\hat{i}1}, ..., \beta_{\hat{i}k})$, k is the number of properties of a profile;

$\varepsilon_{\hat{i}j}$: i.i.d norm error term $Norm(0, \sigma^2)$;

Γ : matrix of coefficients that relates $\beta_{\hat{i}}$ to the value of $z_{\hat{i}}$;

$z_{\hat{i}}$: respondent's information, is a $(3, 1)$ dimension vector;

$\xi_{\hat{i}}$: unobserved heterogeneity component;

$w_{\hat{i}}$: the weight, Gamma distribution with parameters $\nu_1/2, \nu_1/2$.

3.7.1 Gibbs Sampling

Then we draw each parameter using the Gibbs sampling method (Geman and Geman (1984)) according to the posterior distribution mentioned above.

1. Draw $\beta_{\hat{i}}$, $\hat{i} = 1, ..., N$(one respondent at a time).

$$\beta_{\hat{i}} | \{x, y, Z, \Gamma, \sigma^2, \Sigma, w_{\hat{i}}\} \sim Norm(\beta^*_{\hat{i}}, B_{\hat{i}})$$

where

$$B_{\hat{i}} = ((\Sigma/w_{\hat{i}})^{-1} + \sigma^{-2}(\sum_j x_{\hat{i}j}x'_{\hat{i}j}))^{-1},$$

$$\beta^*_{\hat{i}} = B_{\hat{i}}((\Sigma/w_{\hat{i}})^{-1}\Gamma z_{\hat{i}} + \sigma^{-2}(\sum_j x_{\hat{i}j}x'_{\hat{i}j})\hat{\beta}_{\hat{i}}),$$

$$\hat{\beta}_{\hat{i}} = (\sum_j x_{\hat{i}j}x'_{\hat{i}j})^{-1}(\sum_j x_{\hat{i}j}y_{\hat{i}j}).$$

This is the regression result.

2. Draw Γ (Define $\gamma = vec(\Gamma'), \beta = (\beta'_1, ..., \beta'_{\hat{i}})'$).

$$\Gamma | Z, \beta, \Sigma, w \sim Norm(\gamma^*, G)$$

where

$$G = ((Z^*)'(I \otimes (\Sigma/w_{\hat{i}})^{-1})Z^*))^{-1},$$

$$\gamma^* = G((Z^*)'(I \otimes (\Sigma/w_{\hat{i}})^{-1}\beta),$$

$$Z^* = (I \otimes z_1, ..., I \otimes z_3).$$

This is also from the regression result. $\otimes$ is the Kronecker product of two matrixes. $vecA = (a_1, a_2, ..., a_n)'$ is a vec operator of A. A is a $(m \times n)$ matrix, and its j-th column is a_j.

3. Draw w.

$$w | Z, \{\beta_{\hat{i}}\}, \Gamma, \Sigma, \nu \sim \mathcal{G}amma((\nu_1 + 1)/2, (\nu_1 + s)/2)$$

$$s = (\beta - \Gamma Z)\Sigma^{-1}(\beta - \Gamma Z)'$$

4. Draw Σ.

$$\Sigma | \beta, \Gamma, w \sim IW(d_0 + N, D_0 + \sum_{\hat{i}=1}^{N} w_{\hat{i}}(\beta_{\hat{i}} - \Gamma z_{\hat{i}})(\beta_{\hat{i}} - \Gamma z_{\hat{i}})')$$

where prior $\Sigma \sim IW(d_0, D_0)$, and N is the number of respondents.

5. Generate σ^2.

$$\sigma^2|y, x, \{\beta_{\hat{i}}\} \sim Inverted \quad \chi^2(s_0 + n, S_0 + \sum_{\hat{i}} \sum_{j} (y_{\hat{i}} - (x'_{\hat{i}j}\beta_{\hat{i}}))^2)$$

where prior $\sigma^2 \sim Inverted \quad \chi^2(s_0, S_0)$ and n is the number of the observation.

6. Draw $y|n_1, X, \{\beta_{\hat{i}}\}, \sigma^2$ by the Accept-Reject method. n_1 is the number of 1's.

3.7.2 Goodness-Fit Test

To test which model fits these data better, some rules are needed, such as the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the log-marginal density (LMD) criterion. AIC and BIC are general criteria for model selection, while LMD is usually used for Bayesian models, especially for the MCMC draw. AIC is not always consistent with BIC.

The Akaike Information Criterion (AIC) is defined below:

$$AIC = -2\log(\hat{L}) + 2K,$$

where $\hat{L}$ is the estimated likelihood function, and K is the number of free parameter estimated.

The Bayesian Information Criterion (BIC) is defined below.

$$BIC = -2\log(\hat{L}) + K\ln(\bar{N}),$$

where $\bar{N}$ is the sample number. We prefer the model with the smallest AIC or BIC.

The LMD is frequently used in comparing Bayesian models. The model with the highest log-marginal densities (LMD) is most supported by the data. We estimate the LMD as the log of the mean of the likelihood values across iterations $\bar{N}$ of the Gibbs sampler (Allenby et al. (1999), Newton and Raftery (1994)):

$$LMD = \log(\frac{1}{\bar{N}} \sum_{i}^{\bar{N}} p(y_i|\beta_i, \sigma_i)^{-1})^{-1}.$$

We select the model with larger LMD. Some definitions used in the following tables are listed.

a) RHlogit: Robust Hierarchical Logit,

b) RHProbit: Robust Hierarchical Probit,

c) wLMD: LMD is calculated with the likelihood of the whole hierarchical model including the first level (logit/probit) and the second level of robust regression,

d) wAIC: AIC is calculated with the likelihood of the whole hierarchical model include the first level (logit/probit) and the second level of the robust regression,

e) wBIC: BIC is calculated with the likelihood of the whole hierarchical model include the first level (logit/probit) and the second level of the robust regression,

f) hLMD: LMD is calculated with the likelihood of the first level (logit/probit),

g) hAIC: AIC is calculated with the likelihood of the first level (logit/probit),

h) hBIC: BIC is calculated with the likelihood of the first level (logit/probit),

i) rLMD: LMD of the second level of the robust regression,

j)rAIC: AIC of the second level of the robust regression,

k) rBIC: BIC of the second level of the robust regression,

l) fe: The forecast error is computed with the 2838 observations.

3.8 Experiments

Data from Allenby et al. (1995) in which two partial profiles of credit cards are presented to 946 respondents and 14,799 observations are used in our experiment. Each respondent was presented with between 13 and 17 paired comparisons. Thus, this data set has a panel structure. There are two data sets. One is the choice attribute; the other is the demographic attribute. All results are based on the 20,000 MCMC draws. The structure of the hierarchy is shown in Figure (5).

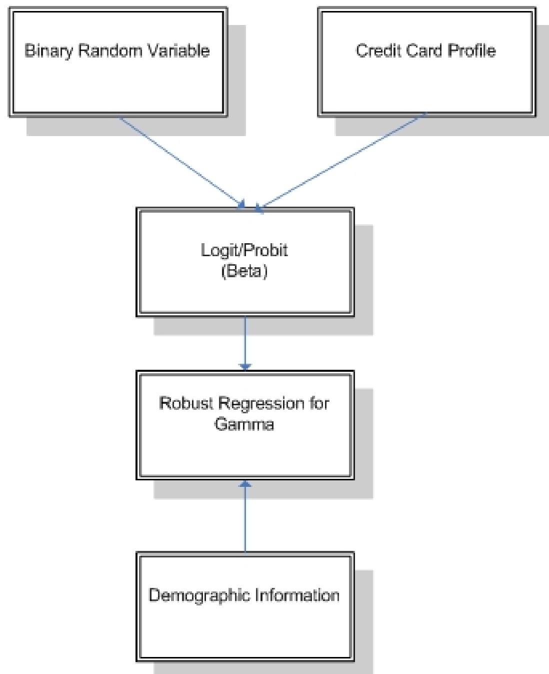

Figure 5: Hierarchical structure.

The outlier analysis, $\beta_{\hat{i}1}$, $\beta_{\hat{i}3}$, and $z_{\hat{i}}$ are sampled for $\beta_{\hat{i}1} = \alpha_{\hat{i}1} + \Gamma_{11} z_{\hat{i}1} + \Gamma_{12} z_{\hat{i}2} + \Gamma_{1,3} z_{\hat{i}3} + \zeta_{(\hat{i},1)}$ and $\beta_{\hat{i}3} = \alpha_{\hat{i}3} + \Gamma_{11} z_{\hat{i}1} + \Gamma_{13} z_{\hat{i}2} + \Gamma_{1,3} z_{\hat{i}3} + \zeta_{(\hat{i},3)}$, Figure (6) shows that outliers do exist from the regression analysis. The data labeled by numbers are assumed to be outliers.

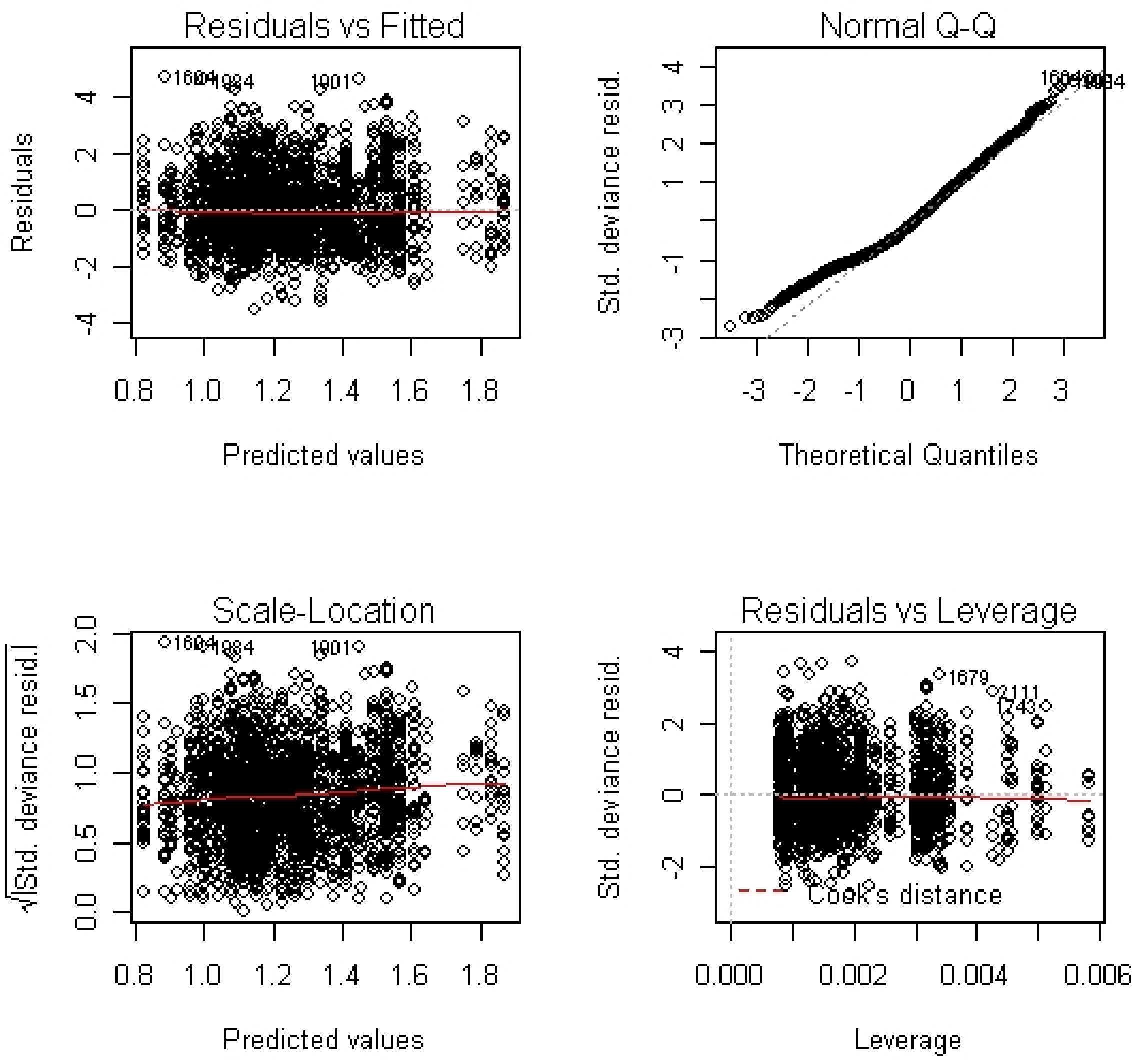

Figure 6: Outlier analysis of $\beta_{\hat{i}1}$, 2000 run.

If we run the simulation 20,000 times, the result will become better, but there are still some outliers as shown in Figure (7). This is motivation for why we use the t distribution.

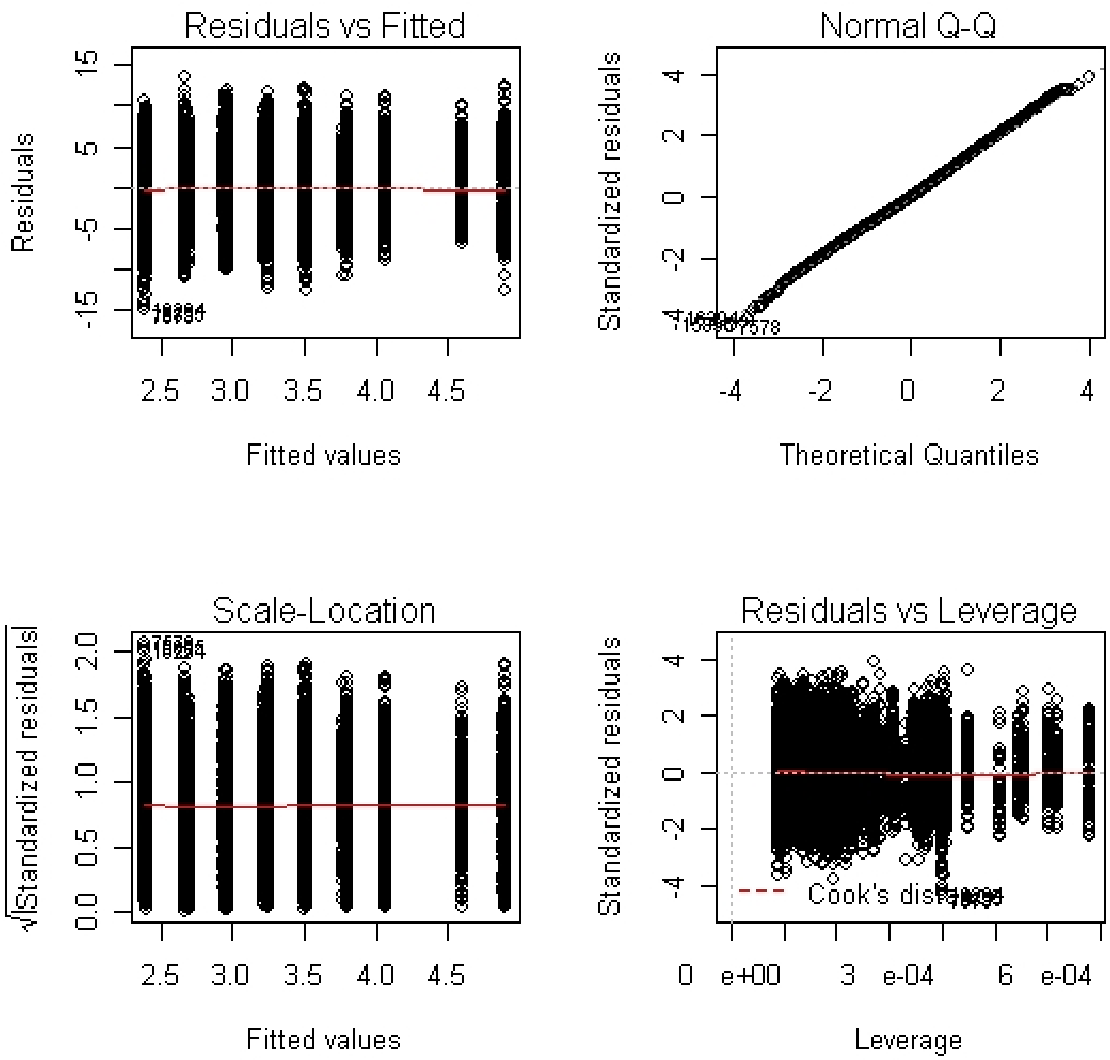

Figure 7: Outlier analysis of $\beta_{\hat{i}3}$, 20,000 run.

In the following likelihood function computation, we do not use the $\hat{i}$ because each response from any respondent is one observation. There is no need to recognize the response from which respondent. The likelihood function of the hierarchical model is

$$L(\beta, \sigma; w, y, z) = \Pi_{i=1}^{n} exp[-\sum_{i=1}^{n} w_i(\beta_i - z_i^{'}\Gamma)^2/2\sigma^2]p(y_i|\beta_i).$$

The likelihood in logit/probit level (top level) is

$$prob_i = \exp{(x'_{ij}\beta_i)}/(1+exp(x'_{ij}\beta_i)) \quad or \quad prob = \Phi(x'_{ij}\beta_i),$$
$$L_i = prob_i y_i + (1-prob_i)(1-y_i),$$
$$L_h = \sum_i log(L_i).$$

The likelihood in robust regression level (bottom level) is

$$L_r = \sum_i -0.5w_k(\beta_i - z_i\Gamma)\Sigma^{-1}(\beta_i-(z_i\Gamma))'.$$

The likelihood for the whole hierarchical model, which includes the top and bottom level), is

$$L_w = \sum_i -0.5w_i(\beta_i - z_i\Gamma)\Sigma^{-1}(\beta_i-(z_i\Gamma))' + L_h.$$

The following tables and graphs show the results of experiments. It seems that the robust hierarchical logit/probit model converges faster than the hierarchical logit/probit model. We run MCMC 20,000 iterations and keep every 20th sample draw.

In the following tables, $\beta_{(.,1)}$ relates to the attribute of credit card, medium fixed interest. $\beta_{(.,2)}$ relates to the attribute of low fixed interest, $\beta_{(.,3)}$ relates to the attribute of medium variable interest, $\beta_{(.,4)}$ relates the attribute of reward program 2 , $\beta_{(.,5)}$ relates to the attribute of reward program 3, $\beta_{(.,6)}$ relates to the attribute of reward program 4, $\beta_{(.,7)}$ relates to the attribute of medium annual fee, $\beta_{(.,8)}$ relates to the attribute of low annual fee, $\beta_{(.,9)}$ relates to the attribute of bank b, $\beta_{(.,10)}$ relates to the attribute of out-of-state bank, $\beta_{(.,11)}$ relates to the attribute of medium rebate, $\beta_{(.,12)}$ relates to the attribute of high rebate, $\beta_{(.,13)}$ relates to the high credit line, $\beta_{(.,14)}$ relates to the long grace period.

3.8.1 Results of Hierarchical Bayesian Logit/Probit Models

Figures (8) and (9) show the MCMC draws of Γ and Σ of HB logit and HB probit. The upper two graphs in each figure are the Γ draws, The bottom two graphs in each figure are the Σ draws. Γ and Σ are defined in Section (3.7.1). The x axis is the number of samples kept. Table (1) shows the mean of Γ, namely, $E[\Gamma]$, the sample variance of $E[\Gamma]$, namely, $Var(E[\Gamma])$, of the hierarchical Bayesian logit and probit model. HB represents the hierarchical Bayesian. In these tables, we do not provide some statistics tests such as t test and p value because those results are simulations. MCMC converges to the expected stationary distribution of these parameters with a very weak condition, namely, irreducible.

Table (2) shows the mean and variance of β of each model such as hierarchical Bayesian logit (HB logit) and hierarchical Bayesian probit (HB probit)

Combining these parameters in Tables (1) and (2), we can analyze the marginal effects of these demographic elements. For example, the increase of the long grace period will decrease the probability of choice probability. An increase of the high credit line will attract a selection of customers.

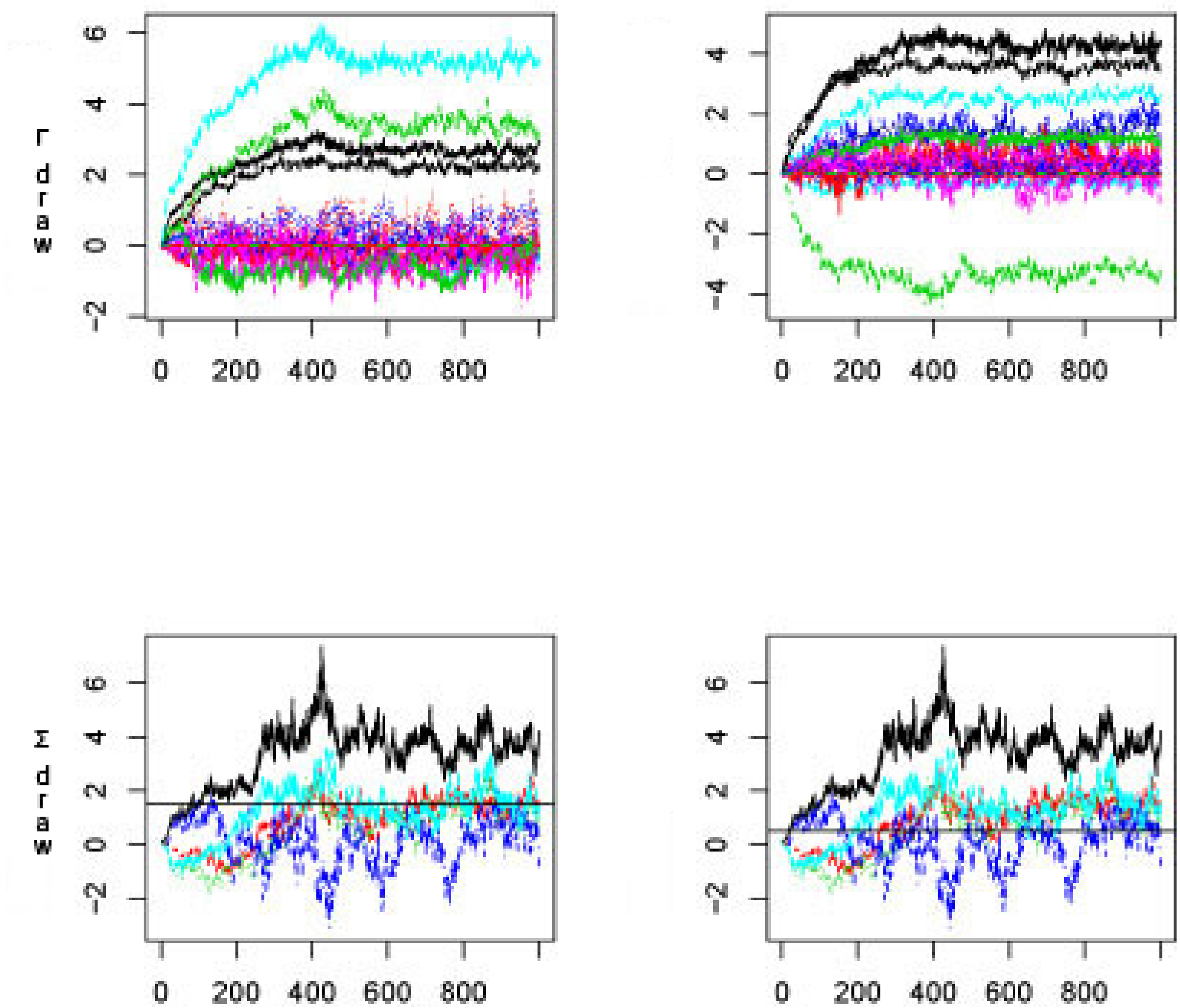

Figure 8: MCMC draw of parameters in hierarchical Bayesian logit.

3.8.2 Results of Robust Hierarchical Bayesian Logit/Probit Model

The following tables and graphs show results of the robust hierarchical Bayesian logit/probit model. Table (3) shows $E[\Gamma]$ and $Var(E[\Gamma])$ of the robust hierarchical Bayesian logit/probit model. Table (4) shows β of each model such as a robust hierarchical Bayesian logit (RHB logit), and robust hierarchical Bayesian probit (RHB probit). Figures (10) and (11) show the MCMC draws of Γ and Σ of RHB logit and RHB probit. The upper two graphs in each figure are the Γ draws; the bottom two graphs in each figure are the Σ draws. Γ and Σ are defined in Section (3.7.1). The x

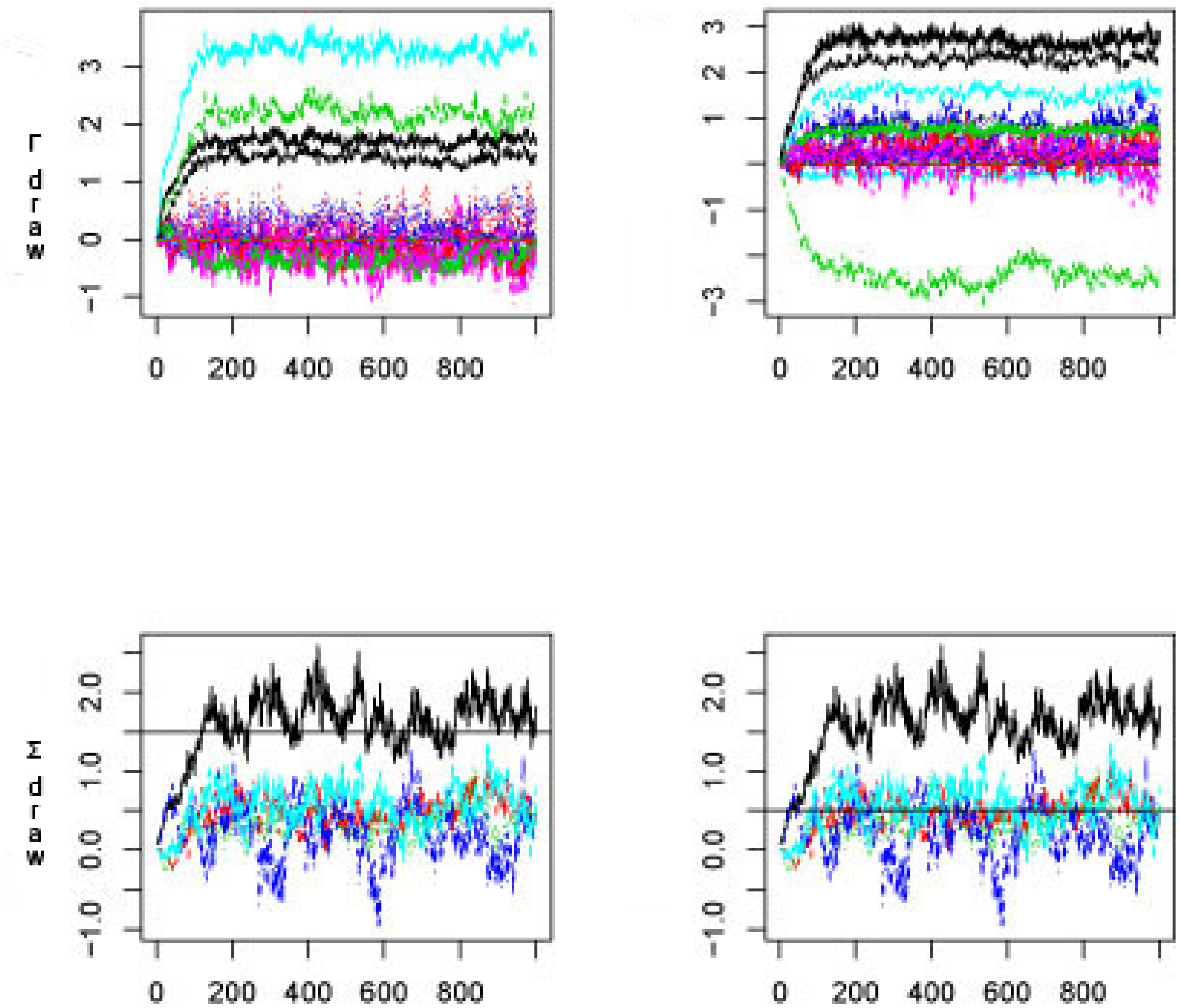

Figure 9: MCMC draw of parameters in hierarchical Bayesian probit.

Table 1: Mean [Γ] and Var[E(Γ)] of hierarchical Bayesian logit and probit.

	E(Γ)							
Attribute	HB Probit				HB Logit			
$\beta(.,.)$	*Intercept*	*age*	*income*	*gender*	*Intercept*	*age*	*income*	*gender*
medium fixed interest	1.699	-0.010	0.008	0.082	2.675	-0.016	0.012	0.181
low fixed interest	3.286	-0.016	0.0150	0.299	5.173	-0.0259	0.022	0.469
medium variable interest	2.145	-0.001	0.018	-0.244	3.403	-0.003	0.025	-0.380
reward program 2	-0.022	0.004	-0.00097	-0.206	-0.040	0.009	-0.001	-0.312
reward program 3	-0.370	0.017	0.007	-0.248	-0.578	0.029	0.011	-0.277
reward program 4	-0.346	0.018	0.008	-0.279	-0.555	0.031	0.013	-0.328
medium annual fee	1.395	-0.0008	0.002	0.395	2.213	-0.003	0.003	0.660
low annual fee	2.676	-0.0032	0.004	0.784	4.271	-0.007	0.007	1.376
bank b	-0.250	-0.0008	0.002	0.096	-0.317	-0.002	0.002	0.129
out-of-state bank	-2.416	-0.009	0.008	0.049	-3.289	-0.013	0.012	0.101
medium rebate	0.889	-0.004	0.003	0.145	1.438	-0.004	0.005	0.273
high rebate	1.575	-0.009	0.015	0.253	2.566	-0.012	0.025	0.513
high credit line	0.737	-0.006	-0.0008	0.260	1.164	-0.011	-0.0002	0.491
long grace period	2.293	-0.016	0.013	0.197	3.575	-0.024	0.019	0.226
	Var[E(Γ)]							
Attribute	HB Probit				HB Logit			
$\beta(.,.)$	*Intercept*	*Age*	*Income*	*Gender*	*Intercept*	*Age*	*Income*	*Gender*
medium fixed interest	6.881e-03	2.924e-05	1.292e-05	2.383e-02	1.274e-02	7.806e-05	2.526e-05	5.452e-02
low fixed interest	1.963e-02	6.805e-05	3.241e-05	6.015e-02	3.516e-02	2.147e-04	7.467e-05	1.606e-01
medium variable interest	1.993e-02	7.549e-05	3.698e-05	6.175e-02	4.153e-02	2.366e-04	7.057e-05	1.703e-01
reward program 2	3.591e-03	1.739e-05	7.752e-06	1.711e-02	1.486e-02	3.912e-05	1.358e-05	2.811e-02
reward program 3	1.018e-02	5.244e-05	1.965e-05	4.479e-02	4.019e-02	1.055e-04	5.695e-05	9.796e-02
reward program 4	1.516e-02	9.521e-05	3.845e-05	1.030e-01	9.087e-02	1.803e-04	8.621e-05	1.881e-01
medium annual fee	6.938e-03	3.313e-05	1.389e-05	2.562e-02	1.285e-02	9.699e-05	3.123e-05	7.298e-02
low annual fee	1.946e-02	1.014e-04	3.693e-05	6.747e-02	3.599e-02	2.301e-04	1.029e-04	2.421e-01
bank b	4.924e-03	2.221e-05	8.630e-06	2.346e-02	7.373e-03	5.632e-05	1.618e-05	4.814e-02
out-of-state bank	2.726e-02	1.132e-04	4.768e-05	9.270e-02	5.358e-02	2.657e-04	8.094e-05	1.826e-01
medium rebate	5.679e-03	2.534e-05	9.494e-06	1.779e-02	1.426e-02	3.915e-05	2.109e-05	2.797e-02
high rebate	1.660e-02	8.429e-05	3.248e-05	7.098e-02	3.192e-02	1.759e-04	8.001e-05	1.075e-01
high credit line	5.057e-03	3.170e-05	1.035e-05	3.079e-02	1.530e-02	7.572e-05	2.455e-05	6.710e-02
long grace period	7.556e-03	4.683e-05	1.988e-05	4.672e-02	3.051e-02	1.134e-04	3.486e-05	8.377e-02

Table 2: Comparison E[β]and Var[β] of hierarchical Bayesian logit/probit.

Attribute	E[β]		Var[β]	
$\beta(.,.)$	*HBlogit*	*HBprobit*	*HBlogit*	*HBprobit*
medium fixed interest	0.015	0.027	0.076	0.057
low fixed interest	0.991	1.133	1.630	1.938
medium variable interest	0.975	0.987	4.575	2.615
reward program 2	-0.813	-0.362	6.300	2.795
reward program 3	1.738	0.996	2.773	1.301
reward program 4	3.222	1.777	9.736	5.108
medium annual fee	-0.380	-0.428	6.149	3.379
low annual fee	3.088	1.893	5.027	2.232
bank b	2.715	1.662	5.603	2.901
out-of-state bank	1.422	0.875	7.909	2.900
medium rebate	-0.744	-0.348	3.637	1.806
high rebate	1.434	0.922	2.719	1.132
high credit line	3.042	1.920	12.913	4.838
long grace period	-0.495	-0.496	5.351	3.162

HB logit: hierarchical bayesian logit model
HB probit:hierarchical bayesian probit model

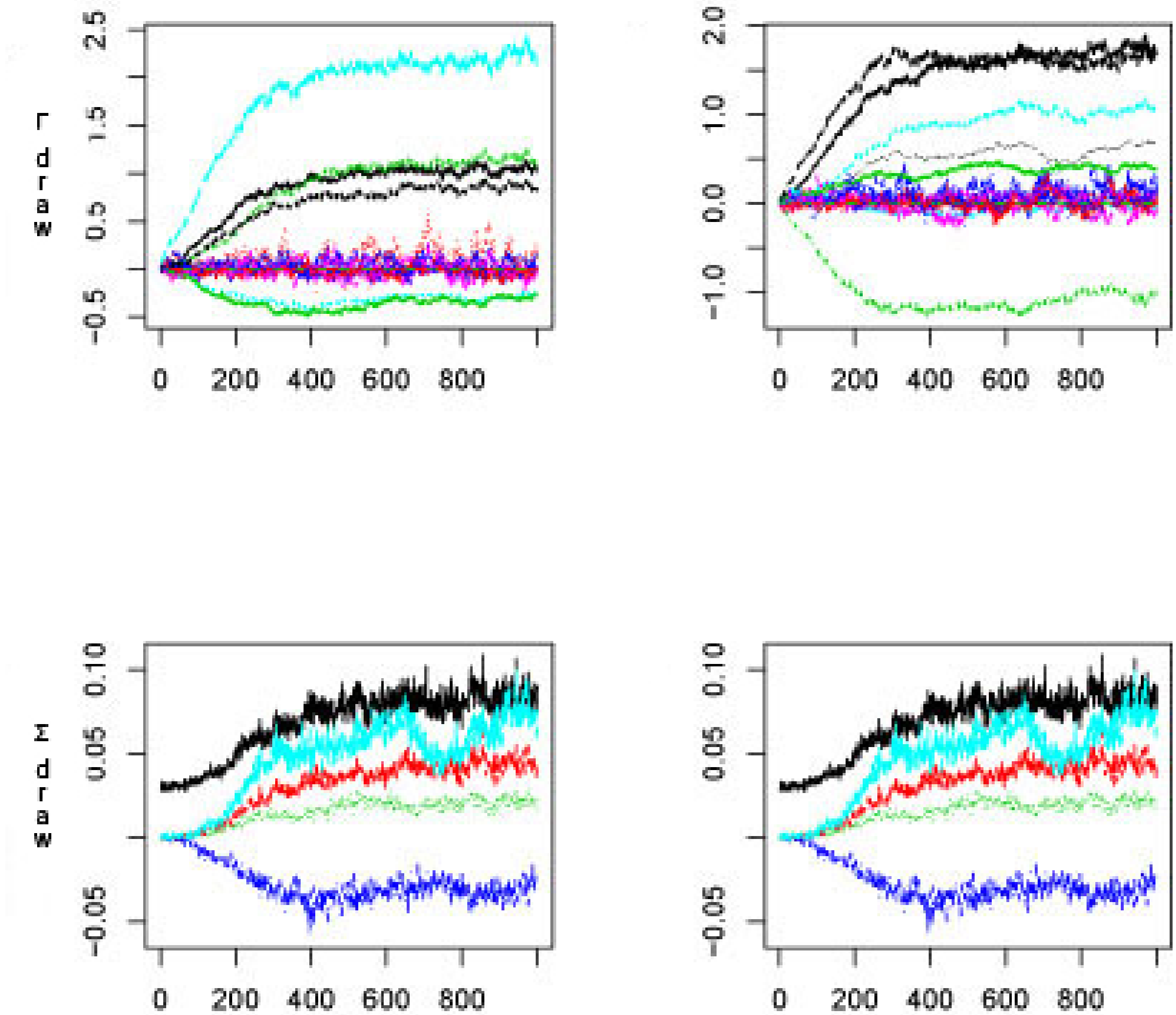

Figure 10: MCMC draw of parameters in robust hierarchical Bayesian logit.

axis is the number of samples kept.

Combining these parameters in Tables (3) and (4), we are able to analyze the marginal effects of these demographic elements. For example, the increase of the long grace period will decrease the probability of choice. An increase of the high credit line will attract a selection chance of customers. The whole trend is the same as the results of the general hierarchical Bayesian logit/probit models.

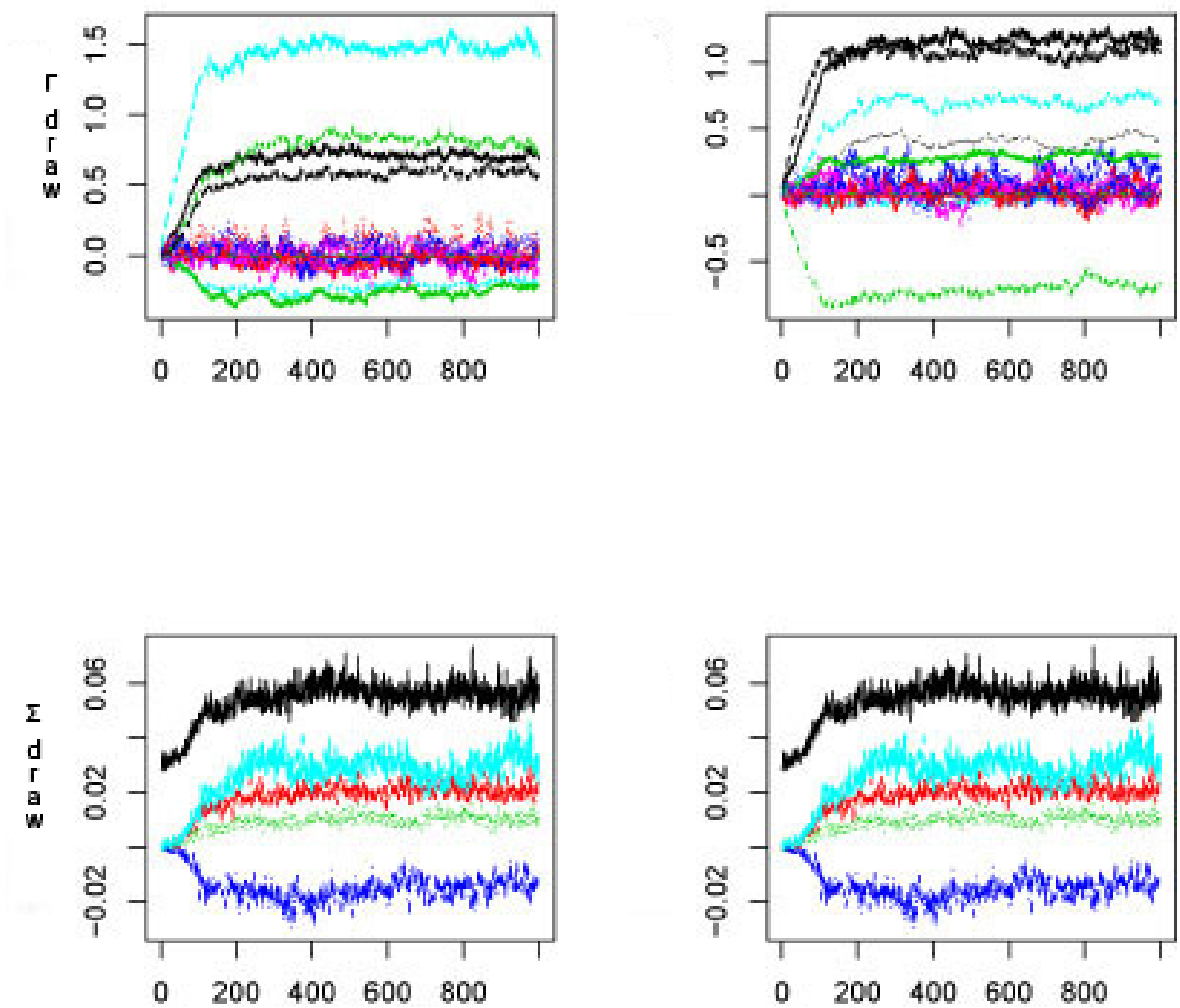

Figure 11: MCMC draw of parameters in robust hierarchical Bayesian probit.

From Tables (5 & 6), the result of the robust hierarchical Bayesian logit/probit model is better than that of the hierarchical Bayesian logit/probit model.

Table 3: $E[\Gamma]$ and $Var[E(\Gamma)]$ of robust hierarchical Bayesian logit and probit.

	E(Γ)							
Attribute	RHB Probit				RHB Logit			
$\beta(.,.)$	*Intercept*	*Age*	*Income*	*Gender*	*Intercept*	*Age*	*Income*	*Gender*
medium fixed interest	0.8951	-0.0053	0.0030	0.0636	1.3511	-0.0066	0.0030	0.0514
low fixed interest	1.8493	-0.0093	0.0062	0.1105	2.7708	-0.0113	0.0064	0.0600
medium variable interest	1.0755	-0.0026	0.0060	-0.0493	1.5507	-0.0035	0.0054	-0.0728
reward program 2	-0.0462	0.0004	-0.0012	-0.0598	-0.0451	-0.0002	-0.0020	-0.0608
reward program 3	-0.2920	0.0043	0.0015	-0.0389	-0.4352	0.0057	0.0012	-0.0155
reward program 4	-0.3596	0.0045	0.0009	0.0007	-0.5321	0.0055	0.0007	0.0306
medium annual fee	0.7924	-0.0050	0.0014	0.1041	1.0891	-0.0050	0.0018	0.0488
low annual fee	1.5525	-0.0073	0.0030	0.2169	2.1432	-0.0077	0.0044	0.1488
bank b	-0.0453	-0.0005	0.0002	0.0311	-0.1275	0.0005	-0.0011	0.0146
out-of-state bank	-0.82230	-0.0014	0.004	0.0305	-1.4119	0.0001	0.0022	-0.0331
medium rebate	0.5693	-0.0022	0.0015	0.0309	0.7308	-0.0016	0.0006	-0.0127
high rebate	0.9949	-0.0063	0.0081	0.0479	1.2628	-0.0067	0.0076	-0.0092
high credit line	0.3719	-0.0042	-0.0005	0.0932	0.5063	-0.0059	-0.0005	0.1314
long grace period	1.2781	-0.0066	0.0055	0.0740	1.9393	-0.0095	0.0059	0.0625
	Var[E(Γ)]							
Attribute	RHB Probit				RHB Logit			
$\beta(.,.)$	*Intercept*	*Age*	*Income*	*Gender*	*Intercept*	*Age*	*Income*	*Gender*
medium fixed interest	8.366e-04	7.737e-06	3.497e-06	5.886e-03	2.629e-03	2.045e-05	8.849e-06	1.888e-02
low fixed interest	2.031e-03	1.088e-05	5.737e-06	1.189e-02	7.566e-03	4.607e-05	2.022e-05	3.238e-02
medium variable interest	1.346e-03	9.347e-06	3.290e-06	7.308e-03	5.749e-03	2.089e-05	1.143e-05	1.706e-02
reward program 2	3.118e-04	2.677e-06	1.093e-06	1.796e-03	5.738e-04	8.574e-06	3.634e-06	3.270e-03
reward program 3	6.045e-04	3.131e-06	1.402e-06	2.769e-03	2.756e-03	9.098e-06	3.565e-06	1.399e-02
reward program 4	1.017e-03	3.539e-06	2.536e-06	5.191e-03	8.123e-03	1.718e-05	1.208e-05	3.952e-02
medium annual fee	6.433e-04	2.279e-06	1.459e-06	3.049e-03	1.960e-03	1.208e-05	8.479e-06	2.061e-02
low annual fee	8.422e-04	6.089e-06	2.182e-06	7.890e-03	5.026e-03	2.923e-05	1.3852e-05	4.301e-02
bank b	9.390e-04	3.053e-06	1.844e-06	4.780e-03	9.469e-04	8.109e-06	4.386e-06	8.804e-03
out-of-state bank	1.986e-03	7.091e-06	4.682e-06	1.203e-02	6.151e-03	3.898e-05	2.134e-05	1.635e-02
medium rebate	4.756e-04	3.695e-06	1.337e-06	2.407e-03	6.611e-03	1.404e-05	4.766e-06	6.568e-03
high rebate	2.509e-03	9.775e-06	3.989e-06	1.104e-02	5.204e-03	2.239e-05	1.069e-05	3.139e-02
high credit line	8.273e-04	3.231e-06	1.039e-06	4.830e-03	1.914e-03	7.903e-06	2.029e-06	1.317e-02
long grace period	1.216e-03	8.428e-06	1.939e-06	7.126e-03	2.867e-03	2.466e-05	9.833e-06	1.834e-02

$\nu of HBlogit = 20, \nu of HBprobit = 20$

Table 4: Comparison E[β] and Var[β] of robust hierarchical Bayesian logit/probit.

Attribute	E[β]		Var[β]	
$\beta(.,.)$	*Rlogit(w)*	*Rprobit(w)*	*Rlogit(w)*	*Rprobit(w)*
medium fixed interest	0.010	0.003	0.045	0.036
low fixed interest	0.335	0.532	0.060	0.069
medium variable interest	0.616	0.422	0.195	0.129
reward program 2	-0.506	-0.317	0.083	0.064
reward program 3	0.897	0.634	0.077	0.076
reward program 4	1.307	0.922	0.327	0.232
medium annual fee	-0.3445	-0.128	0.246	0.119
low annual fee	1.633	1.061	0.440	0.241
bank b	1.219	0.789	0.251	0.191
out-of-state bank	0.687	0.428	0.324	0.171
medium rebate	-0.496	-0.306	0.077	0.065
high rebate	0.826	0.515	0.145	0.095
high credit line	1.620	1.001	0.566	0.315
long grace period	-0.235	-0.101	0.249	0.112

Rlogit(w): robust hierarchical bayesian logit model, with $\nu = 20$
Rprobit(w): robust hierarchical bayesian probit model $\nu = 20$

Table 5: Robust hierarchical Bayesian logit/probit model (500 samples).

Robust Hierarchical Bayesian Logit										
ν	$wLMD$	$wAIC$	$wBIC$	$hLMD$	$hAIC$	$hBIC$	$rLMD$	$rAIC$	$rBIC$	fe
$\star$	-11670.98	11443353	11443825	-5003.416	4843560	4844032	-6841.02	6600017	6600489	-16
7	-10719.91	10534187	10534659	-8126.349	8043093	8043565	-2626.948	2491318	2491790	-2
8	-10805.87	10626944	10627416	-8030.629	7956327	7956799	-2855.613	2670842	2671314	-1
9	-10868.54	10733902	10734374	-7942.388	7894981	7895453	-2980.291	2839145	2839617	-4
10	-11037.44	10855158	10855630	-7931.634	7862585	7863057	-3163.655	2992797	2993269	-3
11	-11156.05	10958913	10959385	-7902.965	7823228	7823700	-3325.279	3135910	3136382	0
12	-11242.28	11065489	11065961	-7873.04	7795967	7796439	-3460.671	3269746	3270218	-3
20	-11958.80	11710985	11711457	-7728.498	7663060	7663532	-4292.588	4048149	4048621	-3
40	-12687.53	12479111	12479583	-7532.458	7450238	7450710	-5259.135	5029097	5029569	-3
Robust Hierarchical Bayesian Probit										
ν	$wLMD$	$wAIC$	$wBIC$	$hLMD$	$hAIC$	$hBIC$	$rLMD$	$rAIC$	$rBIC$	fe
$\star$	-11262.17	10932021	10932493	-4554.749	4381380	4381852	-6803.073	6550865	6551337	-27
7	-10401.11	10234792	10235264	-7822.276	7742459	7742931	-2638.125	2492557	2493029	-5
8	-10508.56	10358401	10358873	-7772.836	7685219	7685691	-2824.978	2673406	2673878	-3
9	-10628.06	10463272	10463745	-7713.193	7622656	7623128	-3021.309	2840840	2841312	-3
10	-10734.51	10568312	10568784	-7648.34	7572063	7572535	-3149.687	2996473	2996945	-2
11	-10818.84	10673712	10674184	-7628.378	7538328	7538800	-3305.076	3135608	3136080	-2
12	-10944.55	10774859	10775331	-7588.046	7508726	7509198	-3408.853	3266357	3266829	-4
20	-11584.86	11384604	11385076	-7430.267	7333211	7333683	-4269.96	4051617	4052089	-5

$\star$ represents hierarchical Bayesian logit/probit model

Table 6: Sensitivity analysis of robust hierarchical Bayesian logit/probit model (500 samples).

Robust Hierarchical Bayesian Logit							
ν	$wLMD$	$wAIC$	$wBIC$	$rLMD$	$rAIC$	$rBIC$	fe
7	0.081	0.079	0.079	0.616	0.623	0.622	0.875
8	0.074	0.071	0.071	0.583	0.595	0.595	0.938
9	0.069	0.062	0.062	0.564	0.570	0.570	0.750
10	0.054	0.051	0.051	0.538	0.547	0.547	0.813
11	0.044	0.042	0.042	0.514	0.525	0.525	1.000
12	0.037	0.033	0.033	0.494	0.505	0.505	0.813
20	-0.025	-0.023	-0.023	0.373	0.387	0.387	0.813
40	-0.087	-0.091	-0.091	0.231	0.238	0.238	0.813
Robust Hierarchical Bayesian Probit							
ν	$wLMD$	$wAIC$	$wBIC$	$rLMD$	$rAIC$	$rBIC$	fe
7	0.076	0.064	0.064	0.612	0.612	0.619	0.815
8	0.067	0.052	0.052	0.585	0.585	0.592	0.889
9	0.056	0.043	0.043	0.556	0.556	0.566	0.889
10	0.047	0.033	0.033	0.537	0.537	0.543	0.926
11	0.039	0.024	0.024	0.514	0.514	0.521	0.926
12	0.028	0.014	0.014	0.499	0.499	0.501	0.852
20	-0.029	-0.041	-0.041	0.372	0.372	0.381	0.815

3.9 Discussion

There are three type of robustness, namely, robustness to prior distribution, data range, and model selections. Our robustness means robustness to prior distribution, and data range. Robustness is not popular because it is difficult to implement at the beginning of its invention. But it has become popular recently because of many new computation techniques such as MCMC.

Bayesian statistics is suitable for a hierarchical model and robust estimation because of its complete condition probability structure. Bayesian regression is better than the least squares (LS) regression. For example, Bayesian regression is naturally a ridge regression overcoming the collinearity difficulty of LS.

MCMC has a very easy satisfying convergence condition, namely, irreducible. The M-H algorithm and Gibbs sampling converge to the stationary distribution of estimated parameters.

Heterogeneity is a basic objective for Market researcher to understand. The market researchers try their best to understand each customer's choice behavior in order to forecast market demand.

A heterogenous model matches the fact that each customer is different. It is supposed to have greater prediction precision because of replacing the same parameters for everyone by specific parameters for each customer.

We also compare the likelihood of the simple logit/probit model with the hierarchical Bayesian logit/probit model, and robust hierarchical Bayesian logit/probit. Our robust models have the highest likelihood; the simple logit/probit has the lowest likelihood.

A hierarchical model is a method for quantifying market heterogeneity. MCMC is an ideal computation technique for a Bayesian statistical model. MCMC is a logical, clear, stable sampling method. The hierarchical logit model seems better than the hierarchical probit model in this case with these data used. A robust hierarchical Bayesian logit/probit model should provide a better fit than a hierarchical Bayesian logit/probit. Our experimental results also demonstrate this advantage.

CHAPTER IV

DYNAMIC PRICING WITH STOCHASTIC DEMAND AND JUMP EVENTS IN PRODUCTION-INVENTORY SYSTEMS

This chapter focuses on the application of brand choice models in dynamic pricing problems. Since McFadden's winning of the Nobel prize in 2000, the logit/probit has found a new application in the dynamic pricing problem.

The objective of dynamic pricing is to maximize revenue by balancing production and inventory, and satisfying customers' demands. Pricing is subject to customers' acceptance. Logit/probit is a well-known brand choice model whose output is the probability of each arriving customer's choosing our products.

In previous literature, most demand processes are assumed to be nonhomogeneous Poisson processes. We assume the rate parameter λ in the nonhomogeneous Poisson process is a stochastic diffusion process. The demand rate process is positive, mean reversion, and has a stable distribution. The Cox-Ingersoll-Ross (CIR) process, a short interest rate process, matches our assumptions.

Production systems often need maintenance and occasional repair. There are many elements affecting production rate. These discrete events are jumps modeled by a mark-time Poisson process.

The final proposed model is a stochastic dynamic programming model. Some concave and monotone structural properties are discussed. Computation is also one of our concerns because of the difficulty of solving the Hamilton-Jacob-Bellman equation in a closed form. The Markov chain approximation method is applied because of its nice convergence property.

Finally, we present some numeric examples in different situations where there are no jumps, jumps with different jump-amplitude distributions, perishable inventory, and logit/probit as customers' response.

We study the following dynamic pricing model. There is a production-inventory system where there are jump events , and customer choices must be considered. Our work considers (1) a complex demand rate λ model, (2) a dynamic pricing model integrating production, inventory, and market information together in the stochastic dynamic programming framework, (3) discussion of structural properties of this type of problem, and (4) numerical examples.

4.1 Introduction

Revenue management began with the reformation of the U.S. airline industry in the late 1970s. Dynamic pricing is a revenue management practice that has spread beyond airlines to the health care, rental car, cruise line, railway, energy, and broadcasting industries.

In recent years, dynamic pricing has attracted the attention of both researchers and entrepreneurs. In particular, there are increased applications in retail industries. Determining the maximum price versus changing demand flow is a two-edged sword.

It can increase revenue, but also rejects the customers. The price is influenced by market information from competitors and present customer loyalty.

A substantial amount of management literature on revenue management (yield management) has been published over the last 20 years. The earliest works on capacity control are Littlewood (1972), Brumelle and McGill (1993), and Curry (1990). Lee and Hersh (1993) introduced and analyzed a discrete-time Markov model that allows for an arbitrary order of arrivals.

In the early stages of dynamic pricing of applications, the goal is to balance supply and demand. This is mainly applied in industries with a constant short-term capacity (supply), such as airlines, cruise ships, hotels, electric utilities, sporting events, and health care (Gallego and van Ryzin (1994, 1997), McGill and van Ryzin (1999), and Weatherford and Bodily (1992)).

In manufacturing systems, Whitin (1955) may have been the first to suggest the need to consider pricing and inventory control strategies together in a non-perishable environment such as retailing. In his paper, Whitin examined a single period problem, very similar to a newsboy problem, and determined a single price and supply quantity. Numerous other researchers have considered price determination and restocking in a multi-period setting. Swann (2001) also considered a dynamic pricing model connected with production-inventory systems whose models are in the deterministic optimization framework.

More recent applications have focused on retail industries, where the short-term supply varies and replenishes, and the object is to optimize inventory management. Advances in the Internet and e-commerce have provided a feasible and reliable method for just-in-time inventory management. A recent survey of revenue management research is provided by McGill and van Ryzin (1999). Barnhart and Talluri (1996) provided an overview of yield management and other airline operations research areas.

The nonhomogeneous compound Poisson process and dynamic programming have important applications in the dynamic pricing problem. Gallego et al. (1997) employed intensity control and obtained structural monotonic results for the optimal intensity (respect to price) as a function of the stock level and the length of the horizon. Bitran et al. (1998) considered a retail chain application. Zhao et al. (2000) considered the given stock of a perishable product over a finite horizon. They also discussed some of the structural properties of this type of dynamic pricing problem.

4.2 Production-Inventory Systems and Brand Choice

4.2.1 Dynamic Pricing in Production-Inventory Systems

The description of a discrete model follows the description by Swann (2001). The mathematical formulation of this pricing problem considers a facility that must determine prices for a single product over a finite horizon. For each period $k \in \{1, ..., T\}$, let X_k, D_k, and I_k be the amount of product produced, the demand satisfied, and inventory at the end of the period k, respectively. This model follows the classic production inventory model.

The system may produce a maximum of Q_k products in the time period k, $R_k(D_k)$ is the revenue function of D_k, and the production cost incurred in period k is c_k per unit produced. Production costs are initially assumed to be linear. Inventory holding cost at a rate of h_k dollars per unit is charged for any inventory carried from period $k-1$ to k. The optimization problem to find the maximum net revenue (profit) and best price, q^*, for each time $k \in \{1, T\}$ with respect to the output vector $X = \{X_k\}$ and price vector $q = \{q_k\}$ is given by the objective

$$\begin{aligned} &\max_{X_k, q_k} \sum_{k=1}^{T} \Big(R_k(D_k) - h_k I_k - c_k X_k \Big) \\ \text{s.t.} \quad & I_{k+1} = I_k + X_k - D_k \\ & X_k \leq Q_k \\ & I_1 = 0, I_k, X_k, D_k \in R^+. \quad k = 1, 2, ..., T \end{aligned} \tag{4.2.1}$$

The object is to find the best price at time $k \in [0, T]$ that makes the maximum revenue during time [0,T]. In our examples, we consider the perishable inventory case.

4.2.2 Combination Logit or Probit with Dynamic Pricing Models

Logit and probit are models for a customer's brand choice. McFadden (1974) completed the analysis: the logit model for the choice probabilities implies that unobserved utility is an extreme value random variable. In his Nobel lecture, McFadden (2001) provided a fascinating history of the development of the model.

Logit and probit models are used to calculate the probability of choice, so the $D_k = logit(q_k\beta) * \lambda_k = \frac{\exp q_k\beta}{1+\exp q_k\beta} * \lambda_k$ or $D_k = probit(q_k\beta) * \lambda_k = \Phi(q_k\beta) * \lambda_k$, where λ_{t_k} is the demand rate for the product and Φ is the std normal cdf. Set $R_k(D_k) = q_k * D_k$ by linearizing; then the original objective function is transformed from

$$\max_{X_k, q_k} \sum_{k=1}^{T} \Big((q_k * D_k) - h_k I_k - c_k X_k \Big)$$

to

$$\max_{X_k,q_k}\sum_{t=1}^{T}\Big(q_k logit(q_k\beta)\lambda_k - h_kI_k - c_kX_k\Big) = \max_{X_k,q_k}\sum_{k=1}^{T}\Big(q_k\frac{\exp q_k\beta}{1+\exp q_k\beta}\lambda_k - h_kI_k - c_kX_k\Big)$$

or

$$\max_{X_k,q_k}\sum_{k=1}^{T}\Big(q_k probit(q_k\beta)\lambda_k - h_kI_k - c_kX_k\Big) = \max_{X_k,q_k}\sum_{k=1}^{T}\Big(q_k\Phi(q_k\beta)\lambda_k - h_kI_k - c_kX_k\Big)$$

This is a nonlinear programming model.

Let the inventory at time k be :

$$I_k = \max[X_k - logit(q_k\beta)\lambda_k, 0] \quad \text{or} \quad I_k = \max[X_k - probit(q_k\beta)\lambda_k, 0]$$

4.3 Stochastic Diffusion Equation of λ of the Demand Poisson Process

Let us begin with some assumptions:

1. Negative demand rates are not allowed.
2. There is a steady distribution for the demand rate.
3. Demand rate has the mean-reverse feature.

We borrow short interest rate models for the dynamic of $\lambda(t)$. There are several short rate models: Rendleman and Bartter model, Vasicek model, and CIR model. We use CIR because it matches our assumptions. Figure (12) shows sample paths of a demand rate $d\lambda(t) = (1/6 - \lambda(t))dt + 0.5\sqrt{\lambda(t)}dW(t)$. $\lambda(t)$ is the demand rate.

Now, we present the continuous-time model. The following are some parameter definitions used in the continuous-time situation. X(t): the production rate. $\lambda(t)$: the

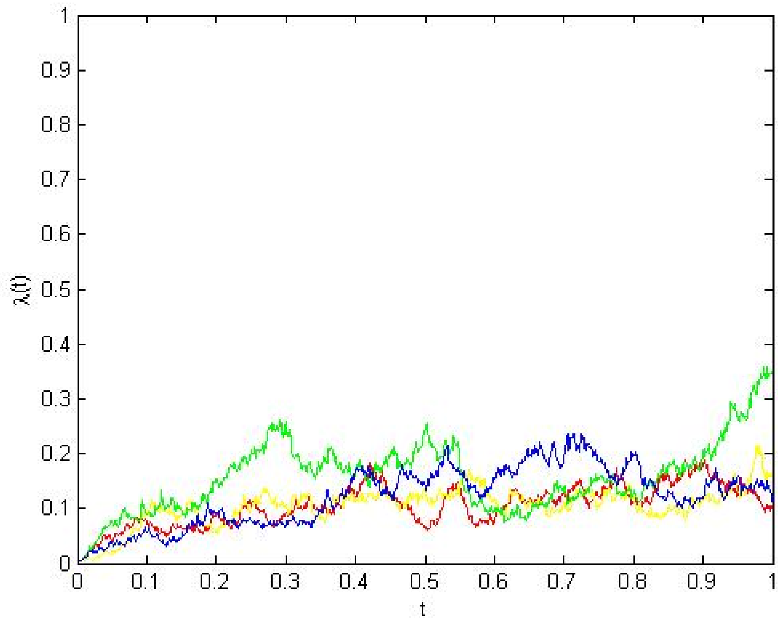

Figure 12: Sample path of CIR process, $\lambda(0) = 0$

demand rate in the market. c(t): the production cost. q(t): the dynamic price. h(t): the inventory cost per time unit. I(t): the inventory rate. In the continuous-time case, we assume the product rate is considered as the decision variable. The models are translated into the following optimal, expected profit model if we use logit as the customer's response function:

$$
\begin{aligned}
V^*(I, X, \lambda, T) &= \max_{q(t)} E \int_0^T \Big\{ q(t) \min[\lambda(t)\bar{H}(q(t))\phi, X(t)] - h(t)I(t) - c(t)X(t) \Big\} dt \\
dX(t) &= f(X(t), \lambda(t), q(t))dt \\
d\lambda(t) &= \alpha \left(\theta - \lambda(t)\right) dt + \sigma \sqrt{\lambda(t)} dW(t) \\
I(t) &= \max[X(t) - \lambda(t)\bar{H}(q(t))\phi, 0] \\
\bar{H}(q(t)) &= logit(q(t)\beta) \quad \text{or} \quad \bar{H}(q(t)) = probit(q(t)\beta) \\
& I(t), X(t), \lambda(t) \in R^+.
\end{aligned}
\tag{4.3.1}
$$

ϕ is the demands at each arrival, α and θ are constants, and σ is the volatility.

Lemma 4.3.1 *If the ϕ is a uniform random variable $unif[a,b]$, $a > 0, b > 0$, and $a \leq \frac{X(t)}{\lambda(t)\bar{H}(q(t))} \leq b$, then*

$$
\begin{aligned}
&V^*(I, X, \lambda, T) \\
&= \max_{q(t)} E \int_0^T \Bigg\{ \frac{q(t)}{b-a}\Bigg((\frac{X(t)^2}{2\lambda(t)\bar{H}(q(t))} - \frac{\lambda(t)\bar{H}(q(t))a^2}{2}) + (bX(t) - \frac{X(t)^2}{\lambda(t)\bar{H}(q(t))})\Bigg) \\
&- \frac{h(t)}{b-a}\Bigg((\frac{X(t)^2}{\lambda(t)\bar{H}(q(t))} - aX(t)) - \frac{1}{2}(\frac{X(t)^2}{\lambda(t)\bar{H}(q(t))} - \lambda(t)\bar{H}(q(t)a^2)\Bigg) - c(t)X(t)\Bigg\} dt
\end{aligned} \tag{4.3.2}
$$

Proof.

If $a \leq \frac{X(t)}{\lambda(t)\bar{H}(q(t))} \leq b$

$$
\begin{aligned}
&E\Bigg[\min[\lambda(t)\bar{H}(q(t))\phi, X(t)]\Bigg] \\
&= \frac{1}{b-a}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t))}} \lambda(t)\bar{H}(q(t))\phi d\phi + \frac{1}{b-a}\int_{\frac{X(t)}{\lambda(t)\bar{H}(q(t))}}^b X(t) d\phi \\
&= \frac{1}{b-a}\Bigg((\frac{X(t)^2}{2\lambda(t)\bar{H}(q(t))} - \frac{\lambda(t)\bar{H}(q(t))a^2}{2}) + (bX(t) - \frac{X(t)^2}{\lambda(t)\bar{H}(q(t))})\Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&E\Bigg[\max[X(t) - \lambda(t)\bar{H}(q(t))\phi, 0]\Bigg] \\
&= \frac{1}{b-a}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t)}} \Bigg(X(t) - \lambda(t)\bar{H}(q(t))\phi\Bigg) d\phi \\
&= \frac{1}{b-a}\Bigg((\frac{X(t)^2}{\lambda(t)\bar{H}(q(t))} - aX(t)) - \frac{1}{2}(\frac{X(t)^2}{\lambda(t)\bar{H}(q(t))} - \lambda(t)\bar{H}(q(t)a^2)\Bigg)
\end{aligned}
$$

If $a \leq b \leq \frac{X(t)}{\lambda(t)\bar{H}(q(t))}$

$$
\begin{aligned}
&E\Big[\min[\lambda(t)\bar{H}(q(t))\phi, X(t)]\Big] \\
&= \frac{1}{b-a}\int_a^b \lambda(t)\bar{H}(q(t))\phi d\phi \\
&= \frac{1}{b-a}\left(\frac{\lambda(t)\bar{H}(q(t))b^2}{2} - \frac{\lambda(t)\bar{H}(q(t))a^2}{2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&E\Big[\max[X(t) - \lambda(t)\bar{H}(q(t))\phi, 0]\Big] \\
&= \frac{1}{b-a}\int_a^b \Big(X(t) - \lambda(t)\bar{H}(q(t))\phi\Big)d\phi \\
&= \frac{1}{b-a}\Big((X(t)b - aX(t)) - \frac{1}{2}\lambda(t)\bar{H}(q(t)(b^2 - a^2)\Big)
\end{aligned}
$$

If $\frac{X(t)}{\lambda(t)\bar{H}(q(t))} \leq a \leq b$

$$
\begin{aligned}
E\Big[\min[\lambda(t)\bar{H}(q(t))\phi, X(t)]\Big] &= X(t) \\
E\Big[\max[X(t) - \lambda(t)\bar{H}(q(t))\phi, 0]\Big] &= 0
\end{aligned}
$$

□

Lemma 4.3.2 *If the ϕ is a normal random variable with $Norm(u, \sigma^2)$, $a \leq \phi \leq b$, $a > 0, b > 0$, Φ is the cumulative distribution function (cdf) of the norm distribution,* *and* $a \leq \frac{X(t)}{\lambda(t)\bar{H}(q(t))} \leq b$, *then*

$$E\Big[\min[\lambda(t)\bar{H}(q(t))\phi, X(t)]\Big] = \frac{1}{\Phi(\frac{b-u}{\sigma}) - \Phi(\frac{a-u}{\sigma})}\Bigg\{-\frac{\lambda(t)\bar{H}(q(t))\sigma}{\sqrt{2\pi}}\Bigg(e^{-\frac{1}{2}(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)} - u}{\sigma})^2} - e^{-\frac{1}{2}(\frac{a-u}{\sigma})^2}\Bigg)$$

$$+ \lambda(t)\bar{H}(q(t))u\Bigg(\Phi(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)} - u}{\sigma}) - \Phi(\frac{a-u}{\sigma})\Bigg) + X(t)\Bigg(\Phi(\frac{b-u}{\sigma}) - \Phi(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)} - u}{\sigma})\Bigg)\Bigg\}$$

$$E\Big[\max[(X(t) - \lambda(t)\bar{H}(q(t))\phi), 0]\Big]$$
$$= \frac{1}{\Phi(\frac{b-u}{\sigma}) - \Phi(\frac{a-u}{\sigma})}\Bigg\{\Bigg(X(t) - u\lambda(t)\bar{H}(q(t))\Bigg)\Bigg(\Phi(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)} - u}{\sigma}) - \Phi(\frac{a-u}{\sigma})\Bigg)$$
$$+ \frac{\lambda(t)\bar{H}(q(t))\sigma}{\sqrt{2\pi}}\Bigg(e^{-\frac{1}{2}(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)} - u}{\sigma})^2} - e^{-\frac{1}{2}(\frac{a-u}{\sigma})^2}\Bigg)\Bigg\} \tag{4.3.3}$$

Proof.

If $a \leq \frac{X(t)}{\lambda(t)\bar{H}(q(t))} \leq b$

$$\Bigg(\Phi(\frac{b-u}{\sigma}) - \Phi(\frac{a-u}{\sigma})\Bigg)E\Big[\min[\lambda(t)\bar{H}(q(t))\phi, X(t)]\Big]$$
$$= \frac{1}{\sigma\sqrt{2\pi}}\int_a^b \min[X(t), \lambda(t)\bar{H}(q(t))\phi]e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$= \frac{1}{\sigma\sqrt{2\pi}}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t))}} \lambda(t)\bar{H}(q(t))\phi e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi + \frac{1}{\sigma\sqrt{2\pi}}\int_{\frac{X(t)}{\lambda(t)\bar{H}(q(t)}}^b X(t)e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$= \frac{1}{\sqrt{2\pi}}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t))}} \lambda(t)\bar{H}(q(t))\frac{(\phi-u)}{\sigma}e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$+ \frac{1}{\sigma\sqrt{2\pi}}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t))}} \lambda(t)\bar{H}(q(t))ue^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi + X(t)\Bigg(\Phi(\frac{b-u}{\sigma}) - \Phi(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t))} - u}{\sigma})\Bigg)$$
$$= -\frac{\lambda(t)\bar{H}(q(t))\sigma}{\sqrt{2\pi}}\Bigg(e^{-\frac{1}{2}(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)} - u}{\sigma})^2} - e^{-\frac{1}{2}(\frac{a-u}{\sigma})^2}\Bigg)$$
$$+ \lambda(t)\bar{H}(q(t))u\Bigg(\Phi(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t))} - u}{\sigma}) - \Phi(\frac{a-u}{\sigma})\Bigg) + X(t)\Bigg(\Phi(\frac{b-u}{\sigma}) - \Phi(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t))} - u}{\sigma})\Bigg)$$

$$\left(\Phi(\frac{b-u}{\sigma})-\Phi(\frac{a-u}{\sigma})\right)E\left[\max[X(t)-\lambda(t)\bar{H}(q(t))\phi,0]\right]$$
$$=\frac{1}{\sigma\sqrt{2\pi}}\int_a^b \max[X(t)-\lambda(t)\bar{H}(q(t))\phi,0]e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$=\frac{1}{\sigma\sqrt{2\pi}}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t))}}\left(X(t)-\lambda(t)\bar{H}(q(t))\phi\right)e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$=\frac{1}{\sigma\sqrt{2\pi}}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t))}}X(t)e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi-\frac{1}{\sigma\sqrt{2\pi}}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t)}}\lambda(t)\bar{H}(q(t))\phi e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$=X(t)\left(\Phi(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)}-u}{\sigma})-\Phi(\frac{a-u}{\sigma})\right)-\frac{1}{\sqrt{2\pi}}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t)}}\lambda(t)\bar{H}(q(t))\frac{(\phi-u)}{\sigma}e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$-\frac{1}{\sigma\sqrt{2\pi}}\int_a^{\frac{X(t)}{\lambda(t)\bar{H}(q(t)}}\lambda(t)\bar{H}(q(t))ue^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$=\left(X(t)-u\lambda(t)\bar{H}(q(t))\right)\left(\Phi(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)}-u}{\sigma})-\Phi(\frac{a-u}{\sigma})\right)$$
$$+\frac{\lambda(t)\bar{H}(q(t))\sigma}{\sqrt{2\pi}}\left(e^{-\frac{1}{2}(\frac{\frac{X(t)}{\lambda(t)\bar{H}(q(t)}-u}{\sigma})^2}-e^{-\frac{1}{2}(\frac{a-u}{\sigma})^2}\right)$$

If $a \leq b \leq \frac{X(t)}{\lambda(t)\bar{H}(q(t))}$

$$\left(\Phi(\frac{b-u}{\sigma})-\Phi(\frac{a-u}{\sigma})\right)E\left[\min[\lambda(t)\bar{H}(q(t))\phi,X(t)]\right]$$
$$=\frac{1}{\sigma\sqrt{2\pi}}\int_a^b \min[X(t),\lambda(t)\bar{H}(q(t))\phi]e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$=\frac{1}{\sigma\sqrt{2\pi}}\int_a^b \lambda(t)\bar{H}(q(t))\phi e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$=\frac{1}{\sqrt{2\pi}}\int_a^b \lambda(t)\bar{H}(q(t))\frac{(\phi-u)}{\sigma}e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi+\frac{1}{\sigma\sqrt{2\pi}}\int_a^b \lambda(t)\bar{H}(q(t))ue^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi$$
$$=-\frac{\lambda(t)\bar{H}(q(t))\sigma}{2\sqrt{2\pi}}\left(e^{-\frac{1}{2}\frac{(b-u)}{\sigma})^2}-e^{-\frac{1}{2}(\frac{a-u}{\sigma})^2}\right)+\lambda(t)\bar{H}(q(t))u\left(\Phi(\frac{b-u}{\sigma})-\Phi(\frac{a-u}{\sigma})\right)$$

$$
\begin{aligned}
&\left(\Phi(\frac{b-u}{\sigma})-\Phi(\frac{a-u}{\sigma})\right)E\Big[\max[X(t)-\lambda(t)\bar{H}(q(t))\phi,0]\Big]\\
&=\frac{1}{\sigma\sqrt{2\pi}}\int_a^b \max[X(t)-\lambda(t)\bar{H}(q(t))\phi,0]e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi\\
&=\frac{1}{\sigma\sqrt{2\pi}}\int_a^b \left(X(t)-\lambda(t)\bar{H}(q(t))\phi\right)e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi\\
&=\frac{1}{\sigma\sqrt{2\pi}}\int_a^b X(t)e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi-\frac{1}{\sigma\sqrt{2\pi}}\int_a^b \lambda(t)\bar{H}(q(t))\phi e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi\\
&=X(t)\left(\Phi(\frac{b-u}{\sigma})-\Phi(\frac{a-u}{\sigma})\right)-\frac{1}{\sqrt{2\pi}}\int_a^b \lambda(t)\bar{H}(q(t))\frac{(\phi-u)}{\sigma}e^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi\\
&-\frac{1}{\sigma\sqrt{2\pi}}\int_a^b \lambda(t)\bar{H}(q(t))ue^{-\frac{(\phi-u)^2}{2\sigma^2}}d\phi\\
&=\left(X(t)-u\lambda(t)\bar{H}(q(t))\right)\left(\Phi(\frac{b-u}{\sigma})-\Phi(\frac{a-u}{\sigma})\right)\\
&+\frac{\lambda(t)\bar{H}(q(t))\sigma}{2\sqrt{2\pi}}\left(e^{-\frac{1}{2}(\frac{b-u}{\sigma})^2}-e^{-\frac{1}{2}(\frac{a-u}{\sigma})^2}\right)
\end{aligned}
$$

If $\frac{X(t)}{\lambda(t)\bar{H}(q(t))}\leq a\leq b$

$$
\begin{aligned}
E\Big[\min[\lambda(t)\bar{H}(q(t))\phi,X(t)]\Big]&=X(t)\\
E\Big[\max[X(t)-\lambda(t)\bar{H}(q(t))\phi,0]\Big]&=0
\end{aligned}
$$

□

4.4 Structural Properties in no-Jump situation

Compound Poisson model: at time zero, the company has stock n to sell, and $s>0$ and the controlled Poisson demand density is $\widehat{\lambda}(s)=\lambda(q(s))=\lambda(s)*\bar{H}(q(s))\phi$. At time s using the pricing policy as optimal price $q(s)$, and the average demand of each arrival is ϕ^e. $\lambda(s)$ is the original demand rate of the Poisson process. The whole problem can be modeled as

$$V^*(I,X,\lambda,t) = \max_{q(s)} E\int_0^t \Big\{q(s)\min[\widehat{\lambda}(s),X(s)] - c(s)X(s) - h(s)\max[X(s)-\widehat{\lambda}(s),0]\Big\}ds$$
$$s.t. \quad \int_0^t \widehat{\lambda}(s)ds \leq \infty \tag{4.4.1}$$

By condition on the arrival of demands, we have

$$V^*(I,X,\lambda,t+dt) = \max_{q(t)}\Bigg\{\lambda(t)dt\bigg(q(t)E\Big[\min[\widehat{\lambda}(s),X(s)]\Big] + V^*(I-Xdt-\phi,X,\lambda,t)\bigg)$$
$$+(1-\lambda(t)dt)V^*(I-Xdt,X,\lambda,t) - c(t)X(t)dt - h(t)E\Big[\max[X(t)-\widehat{\lambda}(t),0)]\Big]dt\Bigg\}$$
$$\frac{\partial V^*(I,X,\lambda,t)}{\partial t} = \max_{q(t)}\Bigg\{\lambda(t)\bigg(q(t)\Big[E\min[\widehat{\lambda}(s),X(s)]\Big] + V^*(I-\phi,X,\lambda,t)\bigg)$$
$$-\lambda(t)V^*(I,X,\lambda,t) - c(t)X(t) - h(t)E\Big[\max[X(t)-\widehat{\lambda}(t),0]\Big]\Bigg\}$$
$$\frac{\partial V^*(I,X,\lambda,t)}{\partial t} = \max_{q(t)}\Bigg\{\lambda(t)\bigg(q(t)E\Big[\min[\widehat{\lambda}(s),X(s)]\Big] - c(t)X(t) - h(t)E\Big[\max[X(t)-\widehat{\lambda(t)},0]\Big]\bigg)$$
$$-\lambda(t)\bigg(V^*(I,X,\lambda,t) - V^*(I-\phi,X,\lambda,t)\bigg)\Bigg\}$$

We set $g(t) = E\Big[q(t)\min[\widehat{\lambda}(s),X(s)] - c(t)X(t) - h(t)\max[X(t)-\widehat{\lambda}(t),0]\Big] > 0.$

We have the following structural properties when there is no inventory cost and no production, but there is initial stock.

Lemma 4.4.1 *When there is no inventory cost and no production, but there is initial stock, $V^*(X,I,\lambda,t)$ and $q(t)$ have the following properties:*

1. *$V^*(X,I,\lambda,t)$ increases in both I and t. See Zhao et al. (2000).*

2. *$V^*(X,I,\lambda,t)$ is concave in I and t when demand is homogenous where there are no inventory cost. See Gallego et al. (1994). Zhao et al. (2000) extended this to the non-homogenous case. $V^*(X,I,\lambda,t)$ is concave in I at fixed t.*

3. $\Delta V^*(X, I, \lambda, t)$ *increases in* t.

Proof.

$$\frac{\partial V^*(I,X,\lambda,t)}{\partial t} = \max_{q(t)}\left\{\lambda(t)\left[g(t) - E\left(V^*(I,X,\lambda,t) - V^*(I-\phi,X,\lambda,t)\right)\right]\right\}$$

$$= \lambda(t)b(q^*, I, X, \lambda, t)$$

$$\textit{where} \quad b(q^*, I, X, \lambda, t) = \max\left[g(t) - E\left(V^*(I,X,\lambda,t) - V^*(I-\phi,X,\lambda,t)\right)\right]$$

$$\frac{\partial \Delta V^*(I,X,\lambda,t)}{\partial t} = \lambda(t)b(q^*, I, X, \lambda, t) - b(q^{'}, I-\Delta I, X, \lambda, t)$$

$$= \lambda(t)\Big(b(q^*, I, X, \lambda, t) - b(q^{'}, I, X, \lambda, t)$$

$$+ b(q^{'}, I, X, \lambda, t) - b(q^{'}, I-\Delta I, X, \lambda, t)\Big)$$

$$= \lambda(t)\Big(b(q^*, I, X, \lambda, t) - b(q^{'}, I, X, \lambda, t)\Big)$$

$$+ \max_{q(t)}\Big\{\lambda(t)\Big[g(t) - E\Big(V^*(I,X,\lambda,t) - V^*(I-\phi,X,\lambda,t)\Big)$$

$$- g(t) + E\Big(V^*(I-\Delta I,X,\lambda,t) - V^*(I-\Delta I-\phi,X,\lambda,t)\Big)\Big]\Big\}$$

$$= \lambda(t)\Big(b(q^*, I, X, \lambda, t) - b(q^{'}, I, X, \lambda, t)\Big)$$

$$+ \max_{q(t)}\Big\{\lambda(t)E\Big(-\Delta V^*(I,X,\lambda,t) + \Delta V^*(I-\Delta I,X,\lambda,t)\Big)\Big\}$$

$$\geq 0$$

$b(q^*, I, X, \lambda, t) - b(q^{'}, I, X, \lambda, t) \geq 0$, $\quad q^*$ *is the optimal solution,*

$\Delta V^*(I-\Delta I, X, \lambda, t) - \Delta(V^*(I, X, \lambda, t) \geq 0$, $\quad V^*$ *is the concave function of* I,$\Delta I > 0$.

I,X,λ,q V^* *are short for* $I(t), X(t), \lambda(t), q(t), V^*(I, X, \lambda, t)$. $\square$

4. $q(t)$ *decreases in* I *for any given* t.

Proof.

$$\begin{aligned}
0 &> b(q^*, I, X, \lambda, t) - b(q^{'}, I, X, \lambda, t) \\
&= g(q^*, t) - g(q^{'}, t) + \Delta V^*(I, X, \lambda, t) \\
&\geq g(q^*, t) - g(q^{'}, t) + \Delta V^*(I + \Delta I, X, \lambda, t) \\
&= b(q^*, I + \Delta I, X, \lambda, t) - b(q^{'}, I + \Delta I, X, \lambda, t)
\end{aligned}$$

□

4.5 Dynamic Pricing using Stochastic Dynamic Programming Theory

In this section, we present the stochastic dynamic programming framework to solve the dynamic pricing problem.

4.5.1 Models and Algorithm

This research focuses on the continuous-time dynamic pricing model with the stochastic dynamic programming framework. The demand process is modeled by a stochastic diffusion process (CIR). Thus, the system with logit model will become

$$\begin{aligned}
V^*(X, \lambda) &= \max_{q(t)} E \int_0^T \left\{ q(t) \min[\lambda(t)\bar{H}(q(t))\phi, X(t)] - h(t)I(t) - c(t)X(t) \right\} dt \\
dX(t) &= f(X(t), \lambda(t), t)dt \\
d\lambda(t) &= \alpha\left(\theta - \lambda(t)\right)dt + \sigma\sqrt{\lambda(t)}dW(t) \\
I(t) &= \max[(X(t) - \lambda(t)\bar{H}(q(t))\phi), 0] \\
\bar{H}(q(t)) &= logit(q(t)\beta)) \quad \text{or} \quad \bar{H}(q(t)) = probit(q(t)\beta)) \\
I_0 &= 0, \quad X(t) \leq Q(t), \quad I(t), X(t), \lambda(t) \in R^+
\end{aligned} \tag{4.5.1}$$

$V^*(X, \lambda)$ means the expected maximum average profit from 0 to T, and the expected maximum average profit is the function of X, λ.

The HJB equation of this system is:

$$0 = \frac{\partial V^*}{\partial t} + \max_{q(t)} \Bigg\{ q(t) E \Big[\min[\lambda(t)\phi\bar{H}(q(t)), X(t)] - h(t)I(t) \Big] - c(t)X(t) + \frac{1}{2}\sigma^2\lambda(t)\frac{\partial^2 V^*}{\partial \lambda^2}$$
$$+ \alpha(\theta - \lambda(t))\frac{\partial V^*}{\partial \lambda} + \frac{\partial V^*}{\partial X} f(X(t), \lambda(t), q(t)) \Bigg\} \tag{4.5.2}$$

And further, equation (4.5.2) is a forward equation. The HJB equation for this system in discrete time by forward Euler approximation

$$V^*_{k+1} = V^*_k + \max_{q_k} \Bigg\{ E \Big[q_k \min[\lambda_k \phi \bar{H}(q_k), X_k] - h_k I_k \Big] - c_k X_k + \frac{1}{2}\sigma^2 \lambda_k \frac{\partial^2 V^*_k}{\partial \lambda^2}$$
$$+ \alpha(\theta - \lambda_k)\frac{\partial V^*_k}{\partial \lambda} + \frac{\partial V^*_k}{\partial X} f(X_k, \lambda_k, q_k) \Bigg\}$$

When applying finite difference, we get the approximated Markov chain:

$$\begin{aligned}
V^*_{k+1}(\lambda, X) = \max_{q_k} \Bigg\{ & p_k\{(\lambda, X), (\lambda, X)\} V^*_k(\lambda, X) \\
& + p_k\{(\lambda, X), (\lambda \pm h_\lambda, X)\} V^*_k(\lambda \pm h_\lambda, X) \\
& + p_k\{(\lambda, X), (\lambda, X \pm h_X)\} V^*_k(\lambda, X \pm h_X) + \triangle t_k g_k(\lambda, X, q) \Bigg\} \\
p_k\{(\lambda, X), (\lambda, X)\} &= 1 - \Delta t_k \left(\frac{\sigma^2 \lambda_k}{2h_\lambda^2} + \frac{|b_{1k}|}{h_\lambda} + \frac{|b_{2k}|}{h_X}\right) \\
p_k\{(\lambda, X), (\lambda \pm h_\lambda, X)\} &= \Delta t_k \left\{\frac{\sigma^2 \lambda_k}{2h_\lambda^2} + \frac{b^\pm_{1k}}{h_\lambda}\right\} \\
p_k\{(\lambda, X), (\lambda, X \pm h_X)\} &= \Delta t_k \frac{b^\pm_{2k}}{h_X} \\
\Delta t_k &\le \frac{\sigma^2 \lambda_k}{2h_\lambda^2} + \frac{|b_{1k}|}{h_\lambda} + \frac{|b_{2k}|}{h_X}
\end{aligned} \tag{4.5.3}$$

where

$$g_k(\lambda, X, q) = E\left[\min[\lambda_k \bar{H}(q_k)\phi q_k, X_k] - h_k \max[X_k - \lambda_k \bar{H}(q_k)\phi q_k, 0]\right] - c_k X_k$$

$$b_{1k} = \alpha(\theta - \lambda_k)$$

$$b_{1k}^{+} = \max[b_{1k}, 0], \quad b_{1k}^{-} = -\min[b_{1k}, 0]$$

$$b_{2k} = f_k, \quad b_{2k}^{+} = \max[b_{2k}, 0], \quad b_{2k}^{-} = -\min[b_{2k}, 0]$$

$p_k\{(\lambda, X), (\lambda, X)\}$: transient probability at time k from (λ, X) to (λ, X). $p_k\{(\lambda, X), (\lambda \pm h_\lambda, X)\}$: transient probability at time k from (λ, X) to $(\lambda \pm h_\lambda, X)$. h_λ and h_X are step coefficients of λ and X respectively.

4.5.2 Numerical Examples without Jump Events

We show two numerical examples. Their profits, price, inventory, and production rate are displayed. These examples represent the production process is in the normal state without jump events.

Example one: the logit model is used as the customer's response function. No mark-time Poisson process exists in this example. The total profit, optimal dynamic prices, inventory, and production rate are displayed in Figures (13) and (14). The model is

$$\begin{aligned} V^*(X, \lambda) = \max_{q(t)} E \int_0^T \Big\{ & q(t) \min[\lambda(t)\phi \frac{\exp(-0.2q(t)+5)}{1+\exp(-0.2q(t)+5)}, X(t)] \\ & - 0.5X(t) - 0.2\max[X(t) - \lambda(t)\phi \frac{\exp(-0.2q(t)+5)}{1+\exp(-0.2q(t)+5)}, 0] \Big\} dt \end{aligned}$$

$$d\lambda(t) = 10\,(2 - \lambda(t))\,dt + 0.3\sqrt{\lambda(t)}dW(t)$$

$$X(t) = 20 + sin(2t),$$

$$\phi = unif[1, 40], \quad \lambda(t) \in [2.7, 3.2], \quad X(t) \in [19, 21], \quad t \in [0, 1], \quad T = 1.$$

.

In Figure (13), we can see the profit increases in t and λ. The price increases in λ. In Figure (14), the average inventory decreases in λ.

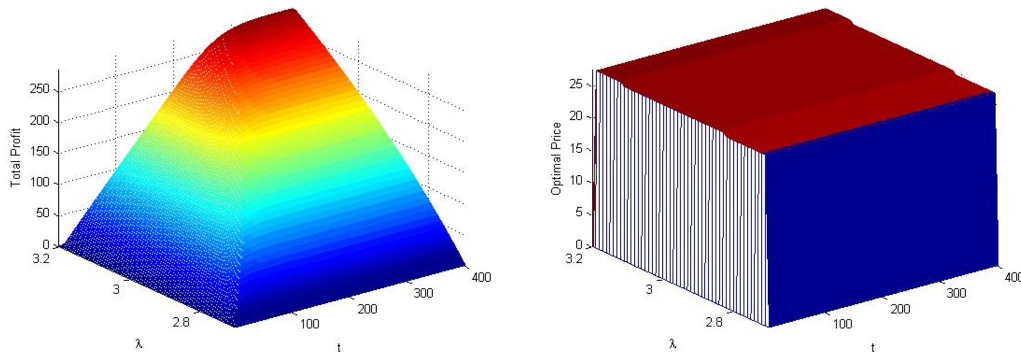

Figure 13: Profit and price in perishable case: no jump, logit.

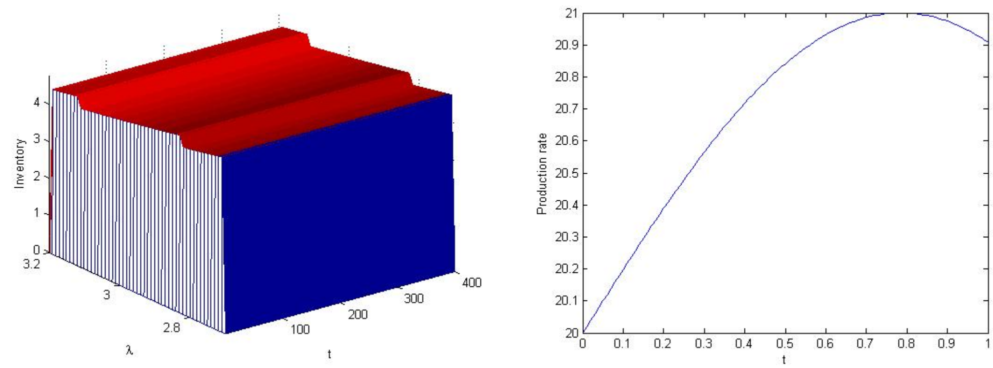

Figure 14: Inventory and production in perishable case: no jump, logit.

Example two: the probit model is used as the customer response function, and without a mark-time Poisson process in the production rate equation. The model is

$$V^*(X,\lambda) = \max_{q(t)} E\int_0^T \Big\{ q(t)\min[\lambda(t)\phi\Phi(-0.2q(t)+5), X(t)] - 0.5X(t)$$
$$- 0.2\max[X(t) - \lambda(t)\phi\Phi(-0.2q(t)+5), 0]\Big\} dt$$
$$d\lambda(t) = 10\,(2-\lambda(t))\,dt + 0.3\sqrt{\lambda(t)}dW(t)$$
$$X(t) = 20 + sin(2t)$$
$$\phi = uinf[1,40], \quad \lambda(t) \in [2.7, 3.2], \quad X(t) \in [19, 21], \quad t \in [0,1], \quad T = 1.$$

The total profit, optimal dynamic prices, inventory, and production rate are displayed in Figures (15) and (16). we can see the profit increases in t and λ. The price increases in λ. The average inventory decreases in λ.

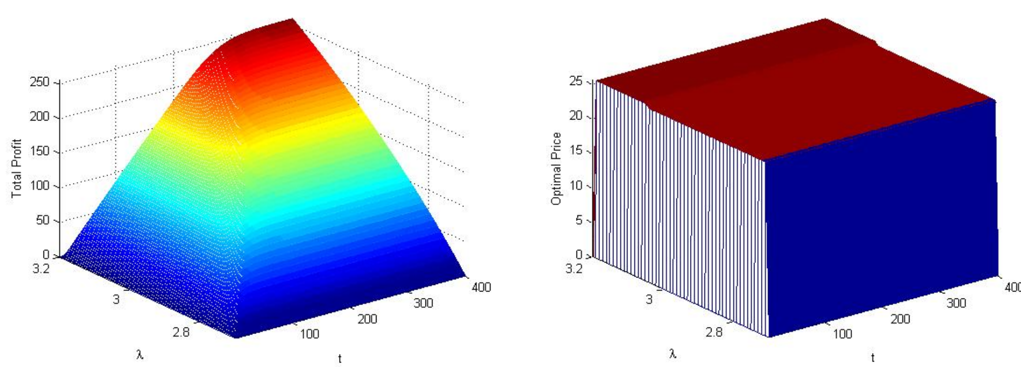

Figure 15: Profit and price in perishable case: no jump, probit.

From the two examples, we can see that the price is increased later in the second example compared with the first example. This relates to the difference between the logit and probit model.

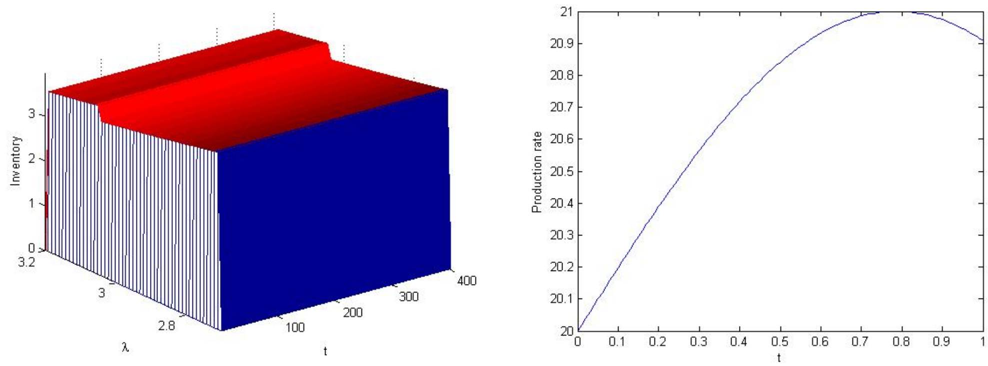

Figure 16: Inventory and production in perishable case: no jump, probit.

4.6 Dynamic Pricing with Jump Events in Production-Inventory Systems

Jump events in production-inventory systems are due to some events such as additional capacity availability, new high technique or additional availability of funds, or external input of this product. Conversely, jump events could include the unavailability of a critical piece of production machinery. $D(t)$ is the demand rate, $I(t)$ is the inventory rate, $X(t)$ is the production rate, $h(t)$ is the inventory cost in unit time, and $c(t)$ is the production cost in unit time. Below are some definitions about jumps:

1. The arrival at time $\tau_1 < \tau_2 < \cdots$, where τ_i is the time of the ith arrival of jump. The arrival rate is $\lambda_J(t)$.

2. The jump amplitude is a random variable drawn from $\mathcal{Q}$ at the time τ; then the time of arrival τ and the jump amplitude $z \in \mathcal{Q}$ from a mark-time Poisson process. The measure of this mark-time Poisson process is

$$E[\mathcal{P}(dz, dt)] = \eta(z) dz \lambda_J(t) dt$$

$\eta(z)$ is the pdf of z, $\lambda_J(t)$ is the rate of the mark-time Poisson process.

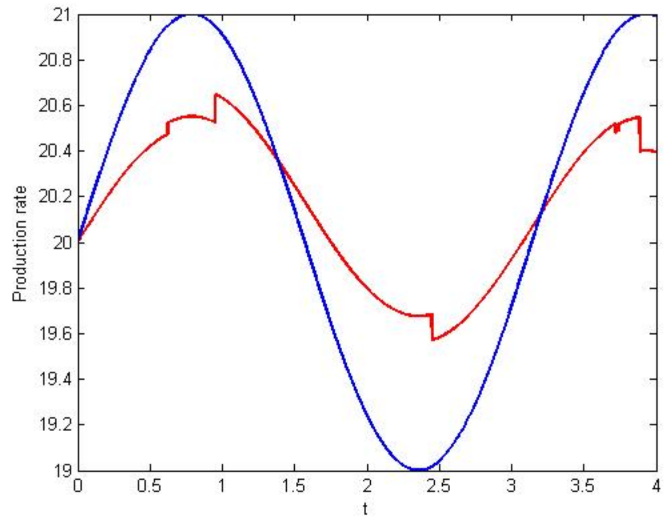

Figure 17: Production rate sample path.

Figure (17) shows a sample path of production rate:

$$dX(t) = cos(2t)dt - dP(t)$$

with jump events whose λ_J=3.0, the amplitude is a uniform random variable [-0.2, 0.2].

The model at this situation is

$$\begin{aligned}
V^*(X, \lambda) &= \max_{q(t)} E \int_0^T \left\{ q(t) \min[\lambda(t)\phi\bar{H}(q(t)), X(t)] - h(t)I(t) - c(t)X(t) \right\} dt \\
dX(t) &= f(X(t), \lambda(t), q(t))dt + dP(t) \\
d\lambda(t) &= \alpha\left(\theta - \lambda(t)\right)dt + \sigma\sqrt{\lambda(t)}dW(t) \\
I_0 &= 0, \quad X(t) \leq Q(t), \quad I(t), \quad X(t), \quad \lambda(t) \in R^+ \\
I(t) &= \max[X(t) - \lambda(t)\phi\bar{H}(q(t)), 0]
\end{aligned} \tag{4.6.1}$$

So we get the HJB equation for the model following the expression of Hanson (2006).

$$
\begin{aligned}
0 = \max_{q(t)} \Bigg\{ & \frac{\partial V^*}{\partial t} + E\Big[\min[\lambda(t)q(t)\phi\bar{H}(q(t)), X(t)] - h(t)I(t)\Big] - c(t)X(t) + \\
& \lambda_J \int_{\mathcal{Q}} (V^*(X+z,t) - V^*(X,t))\eta(z)dz + \frac{\partial V^*}{\partial X} f(X,\lambda,q) \\
& + \frac{1}{2}\sigma^2\lambda(t)\frac{\partial^2 V^*}{\partial \lambda^2} + \alpha(\theta - \lambda(t))\frac{\partial V^*}{\partial \lambda} \Bigg\}
\end{aligned} \tag{4.6.2}
$$

where λ_J is the rate of jumps, and η is the probability density of the jump amplitude.

After constructing the approximated Markov chain from the HJB equation, we get

$$
\begin{aligned}
V^*_{k+1}(X,\lambda) = \max_{q_k} \Bigg\{ & p_k\{(\lambda,X),(\lambda,X)\}V^*_k(X,\lambda) + p_k\{(\lambda,X)(\lambda \pm h_\lambda, X)\}V^*_k(\lambda \pm h_\lambda, X) \\
& + p_k\{(\lambda,X),(\lambda,X \pm h_X)\}V^*_k(\lambda, X \pm h_\lambda) + \triangle t_{k-1} g_k(\lambda,X,q)) + \lambda_J \triangle t \mathcal{J}_k \Bigg\} \\
\mathcal{J}_k = & \int_{\mathcal{Q}} (V^*_k(X+z) - V^*_k(X))\eta(z)dz
\end{aligned} \tag{4.6.3}
$$

where

$$
\begin{aligned}
p_k\{(\lambda,X),(\lambda,X)\} &= 1 - \Delta t_k\left(\frac{\sigma^2\lambda_k}{2h_\lambda^2} + \frac{|b_{1k}|}{h_\lambda} + \frac{|b_{2k}|}{h_X}\right) \\
p_k\{(\lambda,X),(\lambda,X \pm h_X)\} &= \Delta t_k \frac{b^{\pm}_{1k}}{h_X} \\
p_k\{(\lambda,X),(\lambda \pm h_\lambda,X)\} &= \Delta t_k\left\{\frac{\sigma^2\lambda_k}{2h_\lambda^2} + \frac{b^{\pm}_{1k}}{h_\lambda}\right\} \\
\Delta t_k &\leq \frac{\sigma^2\lambda_k}{2h_\lambda^2} + \frac{|b_{1k}|}{h_\lambda} + \frac{|b_{2k}|}{h_X} \\
g_k(\lambda,X,q) &= E\Big[\min[\lambda_k q_k \phi \bar{H}(q_k), X_k] - h_k \max[X_k - \lambda_k q_k \phi \bar{H}(q_k), 0]\Big] - c_k X_k \\
b_{1k} &= f_k, \quad b^+_{1k} = \max[b_{1k}, 0], \quad b^-_{1k} = -\min[b_{1k}, 0] \\
b_{2k} &= \alpha(\theta - \lambda_k) \quad b^+_{2k} = \max[b_{2k}, 0], \quad b^-_{2k} = -\min[b_{2k}, 0]
\end{aligned} \tag{4.6.4}
$$

where z is the jump amplitude value, and h_λ and h_X are step coefficients of λ and X, respectively. $\mathcal{Q}$ is the jump amplitude domain.

4.6.1 Structural Properties in Jump Situation

We have the following structural properties, which are different from the properties without jumps.

Lemma 4.6.1 *1. $V^*(X, I, \lambda, t)$ may not increase in t.*

2. $V^*(X, I, \lambda, t)$ may not be concave in t.

3. $\Delta V^*(X, I, \lambda, t)$ may not increase in t.

Proof.

From equation (4.6.3), we have

$$
\begin{aligned}
V_{k+1}^*(X,\lambda) = \max_{q_k}\Big\{ & V_k^*(X,\lambda) - (1-p_k\{(\lambda,X),(\lambda,X)\})V_k^*(X,\lambda) \\
& + p_k\{(\lambda,X)(\lambda \pm h_\lambda, X)\}V_k^*(\lambda \pm h_\lambda, X) \\
& + p_k\{(\lambda,X),(\lambda,X \pm h_X)\}V_k^*(\lambda, X \pm h_\lambda) + \triangle t_{k-1} g_k(\lambda, X, q)) + \lambda_J \triangle t \mathcal{J}_k \Big\} \\
\mathcal{J}_k = & \int_{\mathcal{Q}} V_k^*(X+z)\eta(z)dz - V_k^*(X)
\end{aligned}
\tag{4.6.5}
$$

If $\mathcal{J}_k = \int_{\mathcal{Q}} V_k^*(X+z)\eta(z)dz - V_k^*(X) < 0$, and

$$
\begin{aligned}
& -(1-p_k\{(\lambda,X),(\lambda,X)\})V_k^*(X,\lambda) + p_k\{(\lambda,X)(\lambda \pm h_\lambda, X)\}V_k^*(\lambda \pm h_\lambda, X) \\
& + p_k\{(\lambda,X),(\lambda,X \pm h_X)\}V_k^*(\lambda, X \pm h_\lambda) + \triangle t_{k-1} g_k(\lambda, X, q)) + \lambda_J \triangle t \mathcal{J}_k < 0
\end{aligned}
\tag{4.6.6}
$$

then $V_{k+1}^*(X,\lambda) < V_k^*(X,\lambda)$. So the $V^*(X,\lambda)$ is no increase by t, the $\Delta V^*(X,I,\lambda,t)$ may not be increasing in t, and $V^*(X,\lambda)$ may not be the concave function of t.

□

4.6.2 Numerical Examples with Jump Events

There are four examples in this section. Two examples are about uniform distributed jump amplitude. The other two are about normally distributed jump amplitude.

Example three: the jump amplitude is a uniform random variable, and the logit model is used as the customer response function. Figure (18) shows the profit and price. The model is

$$\begin{aligned} V^*(X,\lambda) = \max_{q(t)} E \int_0^T \Big\{ & q(t)\min[\lambda(t)\phi\frac{\exp(-0.2q(t)+5)}{1+\exp(-0.2q(t)+5)}, X(t)] \\ & - 0.5X(t) - 0.2\max[X(t)-\lambda(t)\phi\frac{\exp(-0.2q(t)+5)}{1+\exp(-0.2q(t)+5)}, 0] \Big\} dt \end{aligned}$$

$$d\lambda(t) = 10\,(2-\lambda(t))\,dt + 0.3\sqrt{\lambda(t)}dW(t)$$

$$dX(t) = 2cos(2t) + dP(t)$$

dP is a mark-time Poisson process whose amplitude is a $unif[-0.5,-0.05]$. The jump rate $\lambda_J = 1 + sin(t*10)$, t is the time, $X(0) = 20$, $\phi = unif[1,40]$, $\lambda(t) \in [2.7, 3.2]$, $t \in [0,1]$, and $T = 1$. In Figure (18), we can see that the profit does not strictly increases in t, but it still increases in λ. The price increases in λ, but the price is changed only one time compared with the no-jump situation, two times.

Example four: the jump amplitude is a truncated normal random variable. The solution is shown in Figure (19). The model is

$$\begin{aligned} V^*(X,\lambda) = \max_{q(t)} E \int_0^T \Big\{ & q(t)\min[\lambda(t)\phi\frac{\exp(-0.2q(t)+5)}{1+\exp(-0.2q(t)+5)}, X(t)] \\ & - 0.5X(t) - 0.2\max[X(t)-\lambda(t)\phi\frac{\exp(-0.2q(t)+5)}{1+\exp(-0.2q(t)+5)}, 0] \Big\} dt \end{aligned}$$

$$d\lambda(t) = 10*(2-\lambda(t))\,dt + 0.3\sqrt{\lambda(t)}dW(t)$$

$$dX(t) = 2cos(2t)dt + dP(t)$$

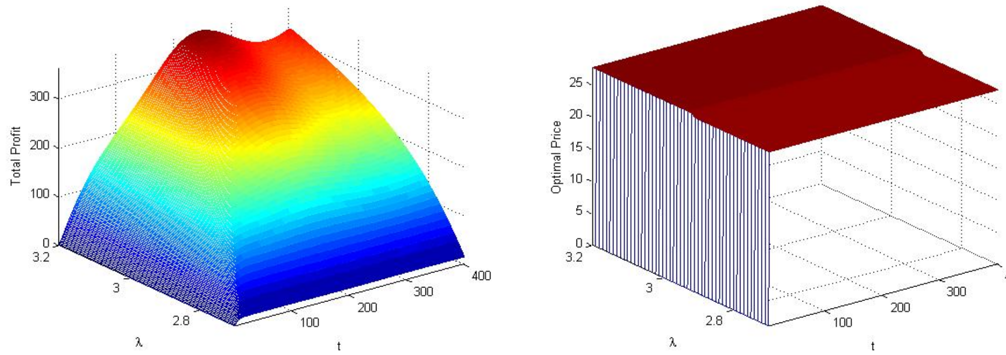

Figure 18: Perishable case: jump amplitude is a unif[-0.5,-0.05], logit.

dP is a mark-time Poisson process whose amplitude is a truncated normal random variable ($[-0.5, -0.05]Norm(0.2, 1/400)$). The jump rate $\lambda_J = 1 + sin(t * 5)$, t is the time, $X(0) = 20$, $\phi = unif[1, 40]$, $\lambda(t) \in [2.7, 3.2]$, $t \in [0, 1]$, and $T = 1$. In Figure (19), we can see that the profit still strictly increases in t and in λ. The price increases in λ, but the price is also less changed compared with the no-jump situation.

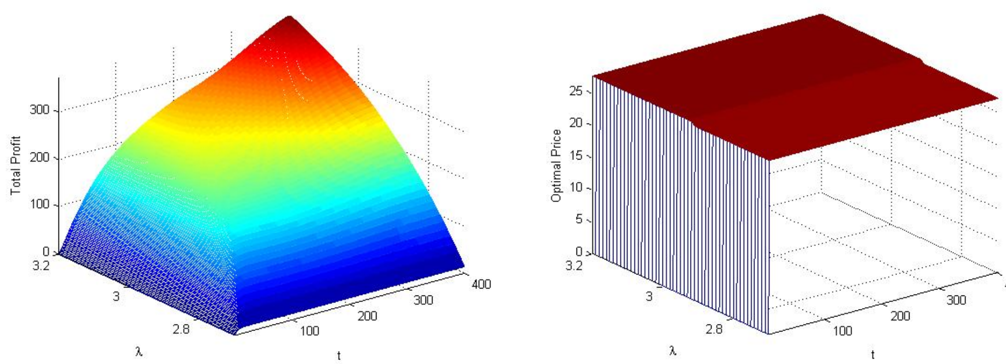

Figure 19: Perishable case: normally distributed jump amplitude, logit.

The following examples are of the probit model used as the customers' response function.

Example five: the jump amplitude is a uniform random variable. Figure (20) is the numerical result.

$$V^*(X,\lambda) = \max_{q(t)} E \int_0^T \Big\{ q(t)\min[\lambda(t)\phi\Phi(-0.2q(t)+5), X(t)] - 0.5X(t)$$
$$- 0.2\max[X(t) - \lambda(t)\phi\Phi(-0.2q(t)+5), 0] \Big\} dt$$

$$d\lambda(t) = 10\,(2-\lambda(t))\,dt + 0.3\sqrt{\lambda(t)}dW(t)$$

$$dX(t) = 2cos(2t)dt + dP(t)$$

dP is a mark-time Poisson process whose amplitude is a $unif[-0.5,-0.05]$. The jump rate $\lambda_J = 1 + sin(t*10)$, t is the time, $X(0) = 20$, $\phi = unif[1,40]$, $\lambda(t) \in [2.7, 3.2]$, $t \in [0,1]$, and $T = 1$.

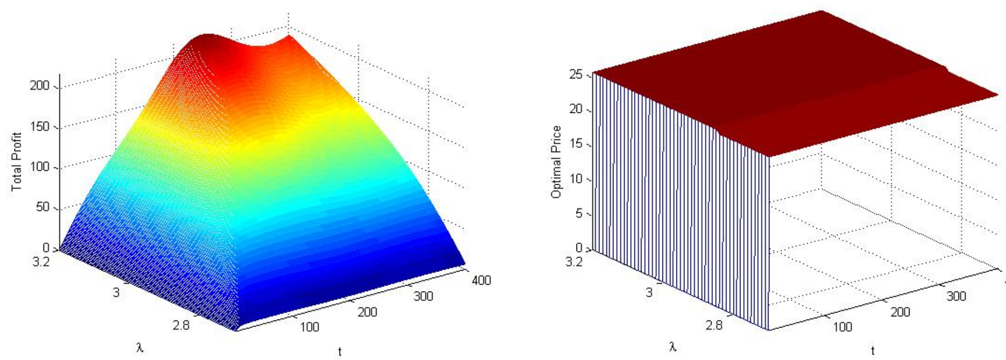

Figure 20: Perishable case: jump amplitude is a unif[-0.5,-0.05], probit.

Example six (Figure 21): the jump amplitude is a truncated normal random variable.

$$V^*(X,\lambda) = \max_{q(t)} E \int_0^T \Big\{ q(t)\min[\lambda(t)\phi\Phi(-0.2q(t)+5), X(t)]$$
$$- 0.5X(t) - 0.2\max[X(t) - \lambda(t)\phi\Phi(-0.2q(t)+5), 0] \Big\} dt$$

$$d\lambda(t) = 10\,(2-\lambda(t))\,dt + 0.3\sqrt{\lambda(t)}dW(t)$$

$$dX(t) = 2cos(2t)dt + dP(t)$$

dP is a mark-time Poisson process whose amplitude is a truncated normal random variable ($[-0.5,-0.05]Norm(0.2, 1/400)$). The jump rate $\lambda_J = 1 + sin(t*5)$, t is the time, $X_0 = 20$, $\phi = unif[1, 40]$, $\lambda(t) \in [2.7, 3.2]$, $t \in [0, 1]$, and $T = 1$. In Figure (21), we can see that the price increases earlier compared with the result of using logit as the customers' response function.

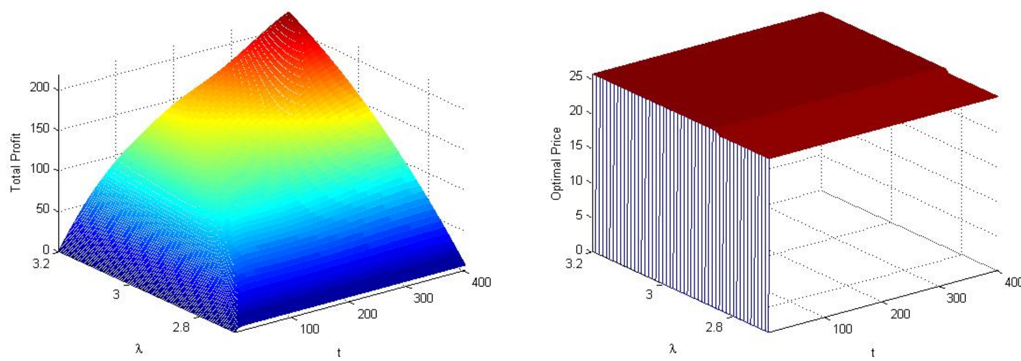

Figure 21: Perishable case: normally distributed jump amplitude, probit.

4.7 Discussion

The objective of dynamic pricing is to increase the profit. But it is also subject to customers' will to buy your products or brands when the price is increased. Sometimes we need to decrease price to clear inventory by attracting customers. The Internet technique makes e-commerce possible and grow fast. Especially dynamic pricing is a characteristic of revenue management based on e-commerce. If e-commerce companies do not utilize the dynamic pricing strategy, they actually waste Internet resources.

The previous dynamic price models seem too simple to use in real situations. For example, inventory costs must be considered. Gallego and Zhao got some structural properties with many assumptions such as no inventory cost and no production. Most of these properties may not hold in a practical situation.

In previous papers, demands are modeled by a simple parameter λ. Actually, λ has its own dynamics, for example, mean-reversion properties. Most supply-demand relations drive the mean-reversion demand of most of products/brands. CIR has a closed-form analysis solution, which is very good for analyzing the demand process.

The ideal situation of a production situation is supposed to be stable. However, there are always some random disturbances, or planned events occur such as maintenance and repair. The mark-time process is a suitable stochastic process for those expected/unexpected events. Actually, Brownian motion is unsuitable for modelling the production process because the diffusion process continuously changes, namely, the frequency of variation of the production process is too high. We do not want the capacity of a production line to change every minute.

Logit and probit have similar predictive precision. The outputs of logit/probit, namely, the customer's choice probability, are close. The function of the logit/probit model represents the customer's response/ feedback to our dynamic price. The feedback structure makes the whole systems stable such that we cannot increase prices at will.

Our model, including the demands model, production model, inventory, and customer's choice, is more practical compared with any previous models. Our model may be able to explore all kinds of situations in real life.

The dynamic price stays the same in most cases, which matches the real situation in a stable market. That means price should not be changed too often; that may not be a good policy.

Homogenous and nonhomogeneous Poisson processes are the theoretical tools to analyze demand processes on which dynamic pricing is based. Dynamic pricing based on a mark-time Poisson process is essentially a stochastic dynamic programming problem , and the dynamic price can be calculated by the Hamilton-Jacobi-Bellman equation. An extended model, which includes the production-inventory information and jump event in this system, is also considered. These numerical results show an insignificant difference in the whole trend of total profit and optimal price between using probit and logit as the customer response function.

CHAPTER V

SUMMARY AND FUTURE WORK

In this dissertation, we developed a new game model, namely, a stochastic differential-jump game model, and applied it to the brand's market competition in a specific situation. Our robust Bayesian hierarchical logit/probit model has more precise estimation and predictive ability. For example, the prediction precision is increased 80% on average compared with the results of a general hierarchical Bayesian logit/probit model in our experiments. Our continuous-time dynamic pricing model is suitable in different situations: jump, or no-jump cases. There are many important opportunities for future research along the direction of our work. We elaborate on these possibilities in the following paragraphs.

(a) The curse of dimensionality of dynamic programming (DP) is also an issue of difficulty when solving the stochastic differential-jump game model developed in this thesis. New computational methods for N-player game model have to be found. Approximate dynamic programming (Powell 2007) is proposed to solve the computational complexity of DP. We also could use an approximate dynamic programming to solve the stochastic differential-jump game model.

(b) We can also apply fuzzy logic and set theory in the market competition model. Fuzzy game (Garagic et al. 2003) can incorporate the players' heuristic knowledge into conventional game theory. Could fuzzy logic be combined with the stochastic differential jump-game framework? If yes, then one must explore strategies for solving the game model.

(c) As with the hierarchical logit/probit model, we can develop a fuzzy regression model (Babuska 1998) in the bottom level. These models may lead to a fuzzy hierarchical logit/probit model. Because logit and probit transform a continuous variable to a discrete variable, their function is similar to the fuzzy characteristic function. Could we develop the robust regression through fuzzy logic?

(d) The variance reduction techniques such as antithetic, control variate, and quasi-Monte Carlo simulation could be employed in a hierarchical model. The hierarchical structure could be complex, and feasible for variance propagation. Antithetic technique is very easy to implement. We sample random elements in opposite directions, and take the average as the mean. Quasi-Monte Carlo simulation is used to find the low difference random number generator. Control variate is to add a correlated random variable to make the estimated variance reduced.

(e) We also could consider a dynamic price model in a nonperishable inventory case where inventory could be accumulated. The inventory cost is increased, and we also may need to set the inventory as an independent state variable to be tracked. The complexity of computation will be increased. In this case, are those structural properties true?

REFERENCES

[1] Allenby, G. M. and Ginter, J. L., "Using extremes to design products and segment markets," *Journal of Marketing Research*, vol. 32, no. 4, pp. 392–403, 1995.

[2] Allenby, G. M. and Lenk, P. J., "Modeling household purchase behavior with logistic normal regression," *Journal of the American Statistical Association*, vol. 89, no. 428, pp. 1218–1231, 1994.

[3] Allenby, G. M., Leone, R. P., and Jen, L., "A dynamic model of purchase timing with application to direct marketing," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 365–374, 1999.

[4] Allenby, G. M. and Rossi, P. E., "Quality perceptions and asymmetric switching between brands," *Marketing Science*, vol. 10, no. 3, pp. 185–204, 1991.

[5] Allenby, G. M. and Rossi, P. E., "Marketing models of customers heterogeneity," *Journal of Econometrics*, vol. 89, no. 1-2, pp. 57–78, 1999.

[6] Anderson, T., *An Introduction to Multivariate Statistical Analysis*. New York: John Wiley and Sons, 1994.

[7] Andrews, R. L. and Currim, I. S., "Behavioral differences between consumers attracted to shopping online vs. traditional supermarkets: Implications for enterprise design and strategy," tech. rep., Department of Business Administration, University of Delaware, Newark, Del., 2000.

[8] Babuska, R., *Fuzzy Modeling for Control*. Boston: Kluwer Academic Publishers, 1998.

[9] Bakos, N. V. and Brynjolfsson, E., "Bundling and competition on the internet," *Marketing Science*, vol. 19, no. 1, pp. 63–82, 2000.

[10] Barnhart, C. and Talluri, K. T., *Airline Operations Research*. New York: John Wiley and Sons, 1996.

[11] Basar, T. and Olsder, G., *Dynamic Noncooperative Game Theory*. London: Academic Press, 1995.

[12] Bass, F., Prasad, A., and Sethi, S., "Generic and brand advertising strategies in a dynamic duopoly," *Marketing Science*, vol. 24, no. 4, pp. 556–568, 2005.

[13] Baumgartner, H. and Homburg, C., "Applications of structural equation modeling in marketing and consumer research: A review," *International Journal of Research in Marketing*, vol. 13, no. 2, pp. 139–161, 1996.

[14] Belk, R. W., "Situational variables and consumer behavior," *The Journal of Consumer Research*, vol. 2, no. 3, pp. 157–164, 1975.

[15] Ben-Akiva and Lerman, S., *Discrete Choice Analysis: Theory and Application to Travel*. Cambridge, Mass.: MIT Press, 1985.

[16] Ben-Akiva, M. and Morikawa, T., "Estimation of switching models from revealed preferences and stated intentions," *Transportation Research*, vol. 24A, no. 6, pp. 485–495, 1990.

[17] Berry, S., Levinsohn, J., and Pakes, A., "Automobile prices in market equilibrium," *Econometrica*, vol. 63, no. 4, pp. 841–890, 1995.

[18] Bitran, G., Caldentey, R., and Mondschein, S., "Coordinating clearance markdown sales of seasonal products in retail chains," *Operations Research*, vol. 46, no. 5, pp. 609–624, 1998.

[19] Blattberg, R. C. and George, E. I., "Shrinkage estimation of price and promotional elasticities: Seemingly unrelated equations," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 304–315, 1991.

[20] Brumelle, S. L. and McGill., J. I., "Airline seat allocation with multiple nested fare classes," *Operations Research*, vol. 41, no. 1, pp. 127–137, 1993.

[21] Brynjolfsson, E. and Smith., M. D., "Frictionless commerce? a comparison of internet and conventional retailers," *Management Science*, vol. 46, no. 4, pp. 563–585, 2000.

[22] Burke, R. R., Harlam, B. A., Kahn, B. E., and Lodish., L. M., "Comparing dynamic consumer choice in real and computer-simulated environments," *The Journal of Consumer Research*, vol. 19, no. 1, pp. 71–82, 1992.

[23] Cardon, J. H. and Hendel, I., "Asymmetric information in health insurance: Evidence from the national medical expenditure survey," *The RAND Journal of Economics*, vol. 32, no. 3, pp. 408–427, 2001.

[24] Cellini, R. and Lambertini, L., "Advertising in a differential oligopoly game," *Journal of Optimization Theory and Applications*, vol. 116, no. 1, pp. 61–81, 2003.

[25] Chintagunta, P. K., Jain, D. C., and Vilcassim, N. J., "Investigating heterogeneity in brand preferences in logit models for panel data," *Journal of Marketing Research*, vol. 28, no. 4, pp. 417–428, 1991.

[26] Currim, I. S., Meyer, R. J., and Le, N. T., "Disaggregate tree-structured modeling of customer choice data," *Journal of Marketing Research*, vol. 25, no. 3, pp. 253–265, 1988.

[27] Curry, R. E., "Optimal airline seat allocation with fare classes nested by origins and destinations," *Transportation Science*, vol. 24, no. 3, pp. 193–204, 1990.

[28] Daganzo, C., *Multinomial Probit: The Theory and its Application to Demand Forecasting*. New York: Academic Press, 1979.

[29] Dhar, S. K., Morrioson, D. G., and Raju, J. S., "The effect of package coupons on brand choice: An epilogue on profits," *Marketing Science*, vol. 15, no. 2, pp. 192–203, 1996.

[30] Degeratu, A. M., Rangaswamy, A., and Wu, J., "Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes," *International Journal of Research in Marketing*, vol. 17, no. 1, pp. 55–78, 2000.

[31] Dickson, P., "Person-situation: Segmentation's missing link," *Journal of Marketing*, vol. 46, no. 4, pp. 56–64, 1982.

[32] Dillon, W. and Gupta, S., "A segment-level model of category volume and brand choice," *Marketing Science*, vol. 15, no. 1, pp. 38–59, 1996.

[33] Dolan, R., Jeuland, A. P., and Muller, E., "Models of new product diffusion: Extension to competition against existing and potential firms over time," Chicago: University of Chicago Center for Research in Marketing, Working paper, 1985.

[34] Eliashberg, J. and Steinberg, R., chapter: Marketing-production joint decision-making, *Handbooks of Operations Research and Management Science: Marketing*, vol. 5, pp. 827–880, Edited by: Eliashberg, J. and Lilien, G. L., North-Holland: Elsevier Science, 1993.

[35] Erdem, T. and Keane, M. P., "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer markets," *Marketing Science*, vol. 15, no. 1, pp. 1–20, 1996.

[36] Erdem, T., Imai, S., and Keane, M. P., "Brand and quantity choice dynamics under price uncertainty," *Quantitative Marketing and Economics*, vol. 1, no. 1, pp. 5–64, 2004.

[37] Erickson, G. M, "Differential game models of advertising competition," *European Journal of Operational Research*, vol. 83, no.3, pp. 431–438, 1995a.

[38] Erickson, G. M, "Advertising strategies in a dynamic oligopoly," *Journal of Marketing Research*, vol. 32, no. 2, pp. 233–237, 1995b.

[39] Esogbue, A., Song, Q., and Young, D., "Non-euler-lagrangian pareto-optimality conditions for dynamic multiple criterion decision problems," *Mathematical Methods of Operations Research*, vol. 63, no. 3, pp. 525–542, 2006.

[40] Feinberg, F. M., Kahn, B. E., and McAlister, L., "Market share response when consumers seek variety," *Journal of Marketing Research*, vol. 29, no. 2, pp. 227-237, 1992.

[41] Feichtinger, G., Hartl, R. F., and Sethi, S. P., "Dynamic optimal control models in advertising: Recent development," *Management Science*, vol. 40, no. 2, pp. 195–226, 1994.

[42] Fennell,G., "Consumer's perception of the product-use situation," *Journal of Marketing*, vol. 42, no. 2, pp. 38–47, 1978.

[43] Feng, Y. and Gallego, G., "Perishable asset revenue management with Markovian time dependent demand intensities," *Management Science*, vol. 46, no. 7, pp. 941–956, 2000.

[44] Garagic, D. and Cruz J. B., "An approach to fuzzy noncooperative Nash games," *Journal of Optimization Theory and Applications*, vol. 118, no. 3, pp. 475–491, 2003.

[45] Gaimon, C., "Simultaneous and dynamic price, production, inventory and capacity decisions," *European Journal of Operational Research*, vol. 35, no. 3, pp. 426–441, 1988.

[46] Gallego, G. and van Ryzin, G., "Optimal dynamic pricing of inventories with stochastic demand over finite horizons," *Management Science*, vol. 40, no. 8, pp. 999–1020, 1994.

[47] Gallego, G. and van Ryzin, G., "A multiproduct dynamic pricing problem and its applications to network yield management," *Operations Research*, vol. 45, no. 1, pp. 24-41, 1997.

[48] Geman, S. and Geman, D., "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, no. 6, pp. 721-741, 1984.

[49] Gensch, D. and Recker, W. W., "The multinational, multiattribute logit choice model," *Journal of Marketing Research*, vol. 16, no. 1, pp. 124–132, 1979.

[50] Geweke, J., "Bayesian treatment of the independent student-t linear model," *Journal of Applied Econometrics, Supplement: Special Issue on Econometric Inference Using Simulation Techniques*, vol. 8, pp. s19–s40, 1993.

[51] Ghosh, A. and McLafferty, S. L., *Location strategies for retail and service firms.* Lexington, Mass.: Lexington Books, 1987.

[52] Gihman, I. I. and Skorohod, A. V., *Stochastic Differential Equations.* New York: Springer-Verlag, 1972.

[53] Goldstein, H. and Rasbash, J., "Improved approximations for multilevel models with binary responses," *Journal of the Royal Statistical Society, A*, vol. 159, no. 3, pp. 505–513, 1996.

[54] Gonul, F. and Shi, M., "Optimal mailing of catalogs: A new methodology using estimable structural dynamic programming models," *Management Science*, vol. 44, no. 9, pp. 1249–1262, 1998.

[55] Gonul, F. and Srinivasan, K., "Modeling multiple sources of heterogeneity in multinomial logit models: Methodological and managerial issues," *Marketing Science*, vol. 12, no. 3, pp. 213–229, 1993.

[56] Gonul, F. and Srinivasan, K., "Estimating the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model," *Marketing Science*, vol. 15, no. 3, pp. 262–280, 1996.

[57] Greene, W., "Sample selection in credit-scoring models," *Japan and the World Economy*, vol. 10, no. 3, pp. 299–316, 1998.

[58] Griliches, Z., "Specification bias in estimates of production functions," *Journal of Farm Economics*, vol. 39, no. 1, pp. 8–20, 1957.

[59] Guadagni, P. M. and Little, J. D. C., "A logit model of brand choice calibrated on scanner data," *Marketing Science*, vol. 26, no. 2, pp. 203–238, 1983.

[60] Gupta, S. and Chintagunta, P. K., "On using demographic variables to determine segment membership in logit mixture models," *Journal of Marketing Research*, vol. 31, no. 1, pp. 128–136, 1994.

[61] Gupta, S., Chintagunta, P., Kaul, A., and Wittink, D. R., "Do household scanner data provide representative inferences from brand choices: A comparison with store data," *Journal of Marketing Research*, vol. 33, no. 4, pp. 383–398, 1996.

[62] Hanson, F. B., *Applied Stochastic Processes and Control for Jump-Diffusions: Modeling, Analysis and Computation.* New York: SIAM, 2006.

[63] Hanson, F. B. and Zhu, Z., chapter: Comparison of market parameters for jump-diffusion distributions using multinomial maximum likelihood estimation, *Proceedings of 43rd IEEE Conference on Decision and Control*, pp. 3919–3924, Edited by: Fradkov, A. L., New York: Springer, 2004.

[64] Hastings, W. K, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97-109, 1970.

[65] Hauser, J. R., Koppelman, F. S., and Tybout, A. M., "Consumer-oriented transportation service planning: Consumer analysis and strategies," *Applications of Management Science*, vol. 1, no. 1, pp. 91–138, 1981.

[66] Hausman, J. and Wise, D., "A conditional probit model for qualitative choice: Discrete decisions recognizing interdependence and heterogeneous preferences," *Econometrica*, vol. 46, no. 2, pp. 403–426, 1978.

[67] Heckman, J. J., "Sample selection bias as a specification error," *Econometrica*, vol. 47, no. 1, pp. 153–161, 1979.

[68] Heckman, J. J., chapter: Statistical model for discrete panel data, *Structural Analysis of Discrete Data and Econometric Applications*, Edited by: Monski, C. and McFadden, D., Cambridge, Mass.: MIT Press, 1981.

[69] Heckman, J. J., "Identifying the hand of past: Distinguishing state dependence from heterogeneity," *The American Economic Reviews*, vol. 81, no. 2, pp. 75–79, 1991.

[70] Heckman, J. and Singer, B., "A method for minimizing the impact of distributional assumptions in econometric models for duration data," *Econometrica*, vol. 52, no. 2, pp. 271–320, 1984.

[71] Helsen, K., Jedidi, K., and DeSarbo, W. S., "A new approach to country segmentation utilizing multinational diffusion patterns," *Journal of Marketing*, vol. 57, no. 4, pp. 60–71, 1993.

[72] Hlavac, T. E. and Little, J. D. C., "A geographic model of an urban automobile market,"*Proceedings of Fourth International Conference on Operations Research*, pp. 302-311, Edited by: Hertz, D. B., New York: Wiley-Interscience, 1969.

[73] Imai, S. and Ching, A., "Bayesian estimation of dynamic discrete choice models," tech. rep., Department of Economics, Queen's University, 2005.

[74] Jeffrey, A. D., "A nested logit model of space and water heat system choice," *Marketing Science*, vol. 5, no. 2, pp. 112–124, 1986.

[75] Jones, J. M. and Landwehr, J. T., "Removing heterogeneity bias from logit models estimation," *Marketing Science*, vol. 7, no. 1, pp. 41–60, 1988.

[76] Jørgensen, S., "A survey of some differential games in advertising," *Journal of Economic Dynamics and Control*, vol. 4, no. 1, pp. 341–369, 1982a.

[77] Jørgensen, S., "A differential games solution to a logarithmic advertising model," *Journal of the Operational Research Society*, vol. 33, no. 5, pp. 425–432, 1982b.

[78] Jørgensen, S., "A differential game solution to an investment and pricing problem," *Journal of the Operational Research Society*, vol. 35, no. 8, pp. 731–739, 1982c.

[79] Jørgensen, S., "Optimal dynamic pricing in an oligopolistic market: A survey," pp. 179–237, *Dynamic Games and Applications in Economics*, Edited by: Basar, T., Berlin: Springer-Verlag, 1986.

[80] Kale, S. H., "Grouping Euroconsumers: A culture-based clustering approach," *Journal of International Marketing*, vol. 3, no. 3, pp. 35–48, 1995.

[81] Kalyanam, K. and Putler, D. S., "Incorporating demographic variables in brand choice models: An indivisible alternatives framework," *Marketing Science*, vol. 16, no. 2, pp. 166–132, 1997.

[82] Kamakura, W. A. and Russell, G. J., "A probabilistic choice model for market segmentation and elasticity structure," *Journal of Marketing Research*, vol. 26, no. 4, pp. 379–390, 1989.

[83] Kamakura, W. A. and Russell, G. J., "Measuring brand value with scanner data," *International Journal of Research in Marketing*, vol. 10, no. 1, pp. 9–22, 1993.

[84] Kamakura, W. A., Ramaswami, S., and Srivastava, R. K., "Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services," *International Journal of Research in Marketing*, vol. 8, no. 4, pp. 329–349, 1991.

[85] Kamakura, W. A., Kim, B.-D., and Lee, J., "Modeling preference and structural heterogeneity in consumer choice," *Marketing Science*, vol. 15, no. 2, pp. 152–172, 1996.

[86] Kannan, P. and Wright, G., "Modeling and testing structured market: A nested logit approach," *Marketing Science*, vol. 10, no. 1, pp. 58–82, 1991.

[87] Krishnamurthi, L. and Raj, S. P., "A model of brand choice and purchase quantity sensitivities," *Marketing Science*, vol. 7, no. 1, pp. 1–20, 1988.

[88] Kushner, H. J., "Numerical approximations for stochastic differential games," *SIAM Journal on Control and Optimization*, vol. 41, no. 2, pp. 457–486, 2002.

[89] Kushner, H. J., "Numerical approximations for stochastic differential games: The ergodic case," *SIAM Journal on Control and Optimization*, vol. 42, no. 6, pp. 1911–1933, 2004.

[90] Kushner, H. J. and Dupuis, P., "Numerical methods for stochastic control problems in continuous time," *SIAM Journal on Control and Optimization*, vol. 28, no. 5, pp. 999–1048, 1990.

[91] Kushner, H. J. and Dupuis, P., *Numerical Methods for Stochastic Control Problems in Continuous Time*. New York: Springer, 2001.

[92] Lanchester, F. W., "Mathematics in warfare," *The World of Mathematics*, vol. 4, pp. 2138–2157, New York: Simon and Schuster, 1956.

[93] Lange, K. L, Little, R. J. A., and Taylor, J. M. G., "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.

[94] Lee, T. C. and Hersh, M., "A model for dynamic airline seat inventory control with multiple seat bookings," *Transportation Science*, vol. 27, no. 3, pp. 252–265, 1993.

[95] Lindley, D. and Smith, A., "Bayes estimates for the linear model," *Journal of the Royal of Statistical Society, B*, vol. 34, no. 1, pp. 1–41, 1972.

[96] Littlewood, K., "Forecasting and control of passengers," *Proceedings of the 12th AGIFORS Symposium*, pp. 95–128, 1972.

[97] Liu, C. and Rubin, D. B., "ML estimation of the multivariate t distribution ," *Statistics Sinia*, vol. 5, no. 1, pp. 19–39, 1995.

[98] Liu, C., "Bayesian robust multivariate linear regression with incomplete data," *Journal of the American Statistical Association*, vol. 91, no. 435, pp. 1219–1227, 1996.

[99] Luce, R. D., *Individual Choice Behavior*. New York: John Wiley and Sons, 1959.

[100] Luce, R. D. and Suppes, P., chapter: Preference, utility, and subjective probability, *Handbook of Mathematical Psychology*, Edited by: Luce, R. D., Bush, R. R., and Galanter, E., New York: John Wiley and Sons, 1965.

[101] Manski, C. F. and Lerman, S. R, "The estimation of choice probabilities from choice based samples," *Econometrica*, vol. 45, no. 8, pp. 1977–1988, 1977.

[102] Mahajan, V., Muller, E., and Bass, F., "New product diffusion models: A review and directions for research," *Journal of Marketing*, vol. 54, no. 1, pp. 1–26, 1990.

[103] Marschak, J., chapter: Binary choice constraints on random utility indications, pp. 312–329, *Stanford Symposium on Mathematical Methods in the Social Science*, Edited by: Arrow, K. J, Stanford, California: Stanford University Press, 1960.

[104] McCarthy, P. S., Kannan, P., Radha, C., and Wright, G. P., "Estimating loyalty and switching with an application to the automobile market," *Management Science*, vol. 38, no. 10, pp. 1371–1393, 1992.

[105] McCulloch, R. and Rossi, P., "An exact likelihood analysis of the multinomial probit model," *Journal of Econometrics*, vol. 64, no. 1–2, pp. 207–240, 1994.

[106] McFadden, D., *Conditional Logit Analysis of Qualitative Choice Behavior*. New York: Academic Press, 1974.

[107] McFadden, D., "The choice theory approach to market research," *Marketing Science*, vol. 5, no. 4, pp. 275-297, 1986.

[108] McFadden, D., "Economic choices," *The American Economic Review*, vol. 91, no. 3, pp. 351–378, 2001.

[109] McGill, J. I. and van Ryzin, G. J., "Revenue management: Research overview and prospects," *Transportation Science*, vol. 33, no. 2, pp. 233–256, 1999.

[110] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E., "Equations of state calculations by fast computing machines," *The Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087-1092, 1953.

[111] Miller, K. E. and Ginter., J. L., "An investigation of situational variation in brand choice behavior," *Journal of Marketing Research*, vol. 16, no. 1, pp. 111–123, 1979.

[112] Moore, W. and Lehmann, D., "A paired comparison nested logit model of individual preference structures," *Journal of Marketing Research*, vol. 26, no. 4, pp. 420-428, 1989.

[113] Moorthy, K. S., "Using game theory to model competition," *Journal of Marketing Research*, vol. 22, no. 3, pp. 262–282, 1985.

[114] Moorthy, K. S., "Theoretical modeling in marketing," *Journal of Marketing*, vol. 57, no. 2, pp. 92–106, 1993.

[115] Morrison, D. G., Massy, W. F., and Silverman, F. N., "The effect of nonhomogeneous populations on Markov steady-state probabilities," *Journal of the American Statistical Association*, vol. 66, no. 334, pp. 268–274, 1971.

[116] Neelamegham, R. and Chintagunta, P., "A Bayesian model to forecast new product performance in domestic and international markets," *Marketing Science*, vol. 18, no. 2, pp. 115–136, 1999.

[117] Newton, M. and Raftery, A. E., "Approximate Bayesian inference by the weighted likelihood bootstrap," *Journal of the Royal Statistical Society, B*, vol. 56, no. 1, pp. 3–48, 1994.

[118] Powell, W. B., *Approximate Dynamic Programming*. New York: John Wiley and Sons, 2007.

[119] Prasad, A. and Sethi, S. P., "Dynamic optimization of an oligopoly model of advertising," tech. rep., Department of Management, University of Texas at Dallas, 2003.

[120] Prasad, A. and Sethi, S. P., "Competitive advertising under uncertainty: A stochastic differential game approach," *Journal of Optimization Theory and Applications*, vol. 123, no. 1, pp. 163–185, 2004.

[121] Punj, G. and Staelin, R., "The choice process for graduate business schools," *Journal of Marketing Research*, vol. 15, no. 4, pp. 588–598, 1978.

[122] Rao, R., chapter: The impact of competition on strategic marketing decision, *The Interface of Marketing and Strategy*, pp. 101–152, Edited by: Wensley, R., Greenwich, Conn.: Jai Press, 1990.

[123] Rao, V., "Price heterogeneity and 'real' inequality: A case study of prices and poverty in rural south India," *The Review of Income and Wealth*, vol. 46, no. 2, pp. 201-211, 2000.

[124] Ratneshwar, S. and Shocker, A., "Substitution in use and the role of usage context in product category structures," *Journal of Marketing Research*, vol. 28, no. 3, pp. 281–295, 1991.

[125] Rju, J. S., Srinivasan, V., and Lai, R., "The effects of brand loyalty on competitive price promotional strategies," *Management Science*, vol. 36, no. 3, pp. 276–304, 1990.

[126] Rodriguez, G. and Goldman, N., "An assessment of estimation procedures for multilevel models with binary responses," *Journal of the Royal Statistical Society, A*, vol. 158, no. 1, pp. 73–89, 1995.

[127] Rossi, P. E. and Allenby, G. M., "A Bayesian approach to estimating household parameters," *Journal of Marketing Research*, vol. 30, no. 2, pp. 171–182, 1993.

[128] Rossi, P. E., Allenby, G. M., and McCulloch, R., *Bayesian Statistics and Marketing*. New York: John Wiley and Sons, 2005.

[129] Rowe, D., *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing*. Boca Raton, Fla.: Chapman & Hall/CRC, 2002.

[130] Roy, R. and Chintagunta, P. K., "A framework for investigating habits, "the hand of the past," and the heterogeneity in dynamic brand choice," *Marketing Science*, vol. 15, no. 3, pp. 280–300, 1996.

[131] Rust, J., "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica*, vol. 55, no. 5, pp. 999–1033, 1987.

[132] Ryan, M. and Gerard, K., "Using discrete choice experiments to value health care programmes: Current practice and future challenges," *Applied Health Economics & Health Policy*, vol. 2, no. 1, pp. 1–10, 2003.

[133] Samuelson, P., "Maximum principles in analytical economics," *Nobel Memorial Lecture*, Singapore: World Scientific, 1970.

[134] Sethi, S. P., "Comparative cluster analysis for world markets," *Journal of Marketing Research*, vol. 8, no. 3, pp. 348–354, 1971.

[135] Sethi, S. P., "Deterministic and stochastic optimization of a dynamic advertising model," *Optimal Control Application and Methods*, vol. 4, no. 2, pp. 179-184, 1983.

[136] Shankar, V., Smith, A. K., and Rangaswamy, A., "Customer satisfaction and loyalty in online and offline environments," *International Journal of Research in Marketing*, vol. 20, no. 2, pp. 153–175, 2003.

[137] Silk, A. J. and Urban, G. L.,"Pre-test-market evaluation of new packaged goods: A model and measurement methodology," *Journal of Marketing Research*, vol. 15, no. 2, pp. 171–191, 1978.

[138] Sorger, G., "Competitive dynamic advertising: A modification of the case game," *Journal of Economics Dynamics and Control*, vol. 13, no. 1, pp. 55–80, 1989.

[139] Stango, V., "Competition and pricing in the credit market," *The Review of Economics and Statistics*, vol. 82, no. 3, pp. 499–508, 2000.

[140] Swann, J., *Dynamic pricing models to improve supply chain performance*. PhD thesis, Department of Industrial Engineering and Management Sciences, Northwestern University, 2001.

[141] Talluri, K. T. and van Ryzin, G. J., "Revenue management under a general discrete choice model of consumer behavior," *Management Science*, vol. 50, no. 1, pp. 15–33, 2004.

[142] Talluri, K. T. and van Ryzin, G. J., *The Theory and Practice of Revenue Management*. New York: Springer, 2006.

[143] Theil, H., *Principles of Econometrics*. New York: John Wiley and Sons, 1971.

[144] Thurstone, L., "The law of comparative judgment," *Psychological Review*, vol. 34, no. 4, pp. 273–286, 1927.

[145] Train, K., "Recreation demand models with taste differences over people," *Land Economics*, vol. 74, no. 2, pp. 230–239, 1998.

[146] Tversky, A. and Kahneman, D., "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, no. 4157, pp. 1124–1131, 1974.

[147] Tversky, A. and Kahneman, D., "The framing of decisions and psychology of choice," *Science*, vol. 211, no. 4481, pp. 453–458, 1981.

[148] Vidale, M. L. and Wolfe, H. B., "An operations-research study of sales response to advertising," *Operations Research*, vol. 5, no. 3, pp. 370–381, 1957.

[149] Von Neumann, J. and Morgenstern, O., *Theory of Games and Economic Behavior*. Princeton: Princeton University Press, 1944.

[150] Weatherford, L. R. and Bodily, S. E., "A taxonomy and research overview of perishable-asset revenue management: Yield management, overbooking, and pricing," *Operations Research*, vol. 40, no. 5, pp. 831–844, 1992.

[151] Wen, C.-H. and Koppelman, F. S., "The generalized nested logit model," *Transportation Research, B*, vol. 35, no. 7, pp. 627–641, 2001.

[152] Whitin, T. M, "Inventory control and price theory," *Management Science*, vol. 2, no. 1, pp. 61–68, 1955.

[153] Yang, S. and Allenby, G. M., "Modeling interdependent consumer preferences," *Journal of Marketing Research*, vol. 40, no. 3, pp. 282–294, 2003.

[154] Yang, S., Allenby, G. M., and Fennell, G., "Modeling variation in brand preference: The roles of objective environment and motivating conditions," *Marketing Science*, vol. 21, no. 1, pp. 14–31, 2002.

[155] Yavas, U., Verhage, B. J., and Green, R. T., "Global consumer segmentation versus local market orientation: Empirical findings," *Management International Review*, vol. 32, no. 3, pp. 265–273, 1992.

[156] Zhao, W., *Dynamic and static yield management models*. PhD thesis, Department of Operations and Information Management, University of Pennsylvania, 1999.